

101 uses for a dead bird

Natural history museums face an unprecedented opportunity as central players in biodiversity research. They and their financial supporters need to match fine words with appropriate actions.

President Thomas Jefferson had a better appreciation of the importance of biodiversity research than many contemporary politicians. When the United States acquired Louisiana Territory from France in the early nineteenth century, he sent scientists to record its climate, biota, minerals and peoples, as well as "the dates at which particular plants put forth or lose their flowers or leaf, times of appearance of particular birds, reptiles or insects". Jefferson's mandate vividly illustrates that natural history collections are much more than aesthetic displays of stuffed birds or dried plants; they are databases compiling taxonomic, systematic and ecological information on species and their habitats.

An invaluable heritage of some 3 billion specimens has accumulated in museums worldwide as a result of expeditions carried out over the centuries. Museum directors, keen to capitalize on contemporary political enthusiasm for biodiversity, are now rightly asserting that their natural history collections are essential to underpin efforts to create the detailed picture of the Earth's biota that will provide a baseline for its responsible management (see page 115). That much has been acknowledged by the signatories to the United Nations Biodiversity Convention, while a strong case for much greater support for museum research is contained in a recent report by the US President's Committee of Advisors on Science and Technology (PCAST).

Fine words from politicians in the United States and elsewhere have yet to be matched by significant increases in funding. But imagine the mess were genome databases of maps and sequences isolated in regional repositories instead of being available to the world's biology community at the click of a Web page. Imagine also that the data had few if any links to related information and lacked reliable quality control, with much being redundant, incomplete or duplicated.

That is the situation that currently reigns in the management and exploitation of many of the world's natural history collections. The most pressing need is for a physical refurbishment to preserve existing collections, along with an international digitization programme to create a seamless web linking taxonomic and systematic information from the world's collections with ecological and biogeographical databases — such as that being proposed by the Megascience Forum of the Organization for Economic Cooperation and Development.

That basic infrastructure is needed if new data are to be efficiently incorporated from efforts to create an inventory of the millions of species that have yet to be described, such as the proposed Systematics Agenda 2000 programme. It is also required in order to put all biodiver-

sity research on a firmer scientific footing. In one sense, biodiversity research is what museums should have been doing for centuries, and their relative unpreparedness suggests that they have fallen short in their role as custodians of the Earth's biodiversity. But museums have often had barely enough funds to keep collections from deteriorating, let alone to begin to think of making them more sophisticated or embarking on vast new inventory efforts.

Nevertheless, museum directors everywhere need to scrutinize their consciences to ensure that they are defending their core missions with sufficient zeal, and not neglecting them through an excessive pursuit of the short-term political favour and revenues that can flow from their public exhibitions. It is also a paradox that whereas biodiversity research is eminently global, museums have often been parochial in outlook and have tended to focus on their own collections. It is increasingly clear that museums can no longer afford to work in isolation.

That message seems to be hitting home. Europe's major museums have formed a consortium, while the PCAST report calls for a coordinated interdisciplinary approach to biodiversity incorporating museums and the disparate federal agencies involved. An embryo of a credible umbrella organization for biodiversity science is also emerging in the form of *Diversitas*, a body under the auspices of UNESCO and the International Council for Scientific Unions.

One priority should be increased scientific collaboration between museums in the diversity-poor but economically rich nations and the diversity-rich but economically poor countries. Not only are developing countries home to most of the world's biodiversity, but the relatively low costs of such research puts excellence within their means. Mexico's National Commission for the Knowledge and Use of Biodiversity provides a shining example of what can be achieved — with an annual budget of just \$2.5 million, it has sent scientists to museums worldwide to digitize existing specimens of Mexico's biota. By integrating the resulting data with ecological and other databases, it has built what many Western scientists acknowledge is one of the most powerful biodiversity management systems around.

The urgent need now is for museums and other bodies involved in biodiversity research to organize and reach international scientific consensus on priorities for funding. Efforts by *Diversitas* and other groups to identify gaps where knowledge is lacking about particularly ecologically important or endangered groups are a good start, as is the PCAST report's agenda. That process of prioritization needs to be accelerated and deepened. □

New policy for structure data

Like other leading journals, *Nature* has in recent months been considering a change in editorial policy, requiring that high-resolution structural coordinate data be made freely available at the time of publication, rather than allowing the option of a one-year hold on such release (see *Nature* 391, 617; 1998). It is clear that there is a significant majority opinion in the community against permitting a one-year hold. Accordingly, *Nature*, simultaneously with *Science*, is changing its policy. Any paper containing new structural data

received on or after 1 October 1998 will not be accepted without an accession number from the Brookhaven Protein Data Bank (PDB) accompanied by an assurance that unrestricted ("layer-1") release will occur at or before the time of publication (see April 1998 PDB Newsletter at www.pdb.bnl.gov/pdb-docs/newsletter.html). We will undertake to notify the PDB ahead of publication to ensure that the data are unlocked on the appropriate date.

Philip Campbell Editor, *Nature*



US public puts faith in science, but still lacks understanding

[WASHINGTON] Public interest in science and technology is at an all-time high in the United States, even though popular understanding of basic scientific concepts remains weak, according to an extensive study for the National Science Foundation (NSF).

But the study, published last week, reports that this understanding is still greater than that in many other countries, and that US faith in the ability of science to do good is markedly greater than that elsewhere.

The study was conducted by Jon Miller of the Chicago Academy of Sciences, and is incorporated into *Science and Engineering Indicators 1998*, the NSF's biennial report on national trends in science. It found that 70 per cent of Americans say they are interested in science and technology; this is the highest number ever recorded.

The number saying they were interested in science was up from 61 per cent in 1992 and 67 per cent in 1995, and now exceeds those professing interest in foreign policy (47 per cent) and economic policy (68 per cent). The number who considered themselves 'well informed' about scientific discoveries and new technologies had also risen sharply over the past five years, the survey found.

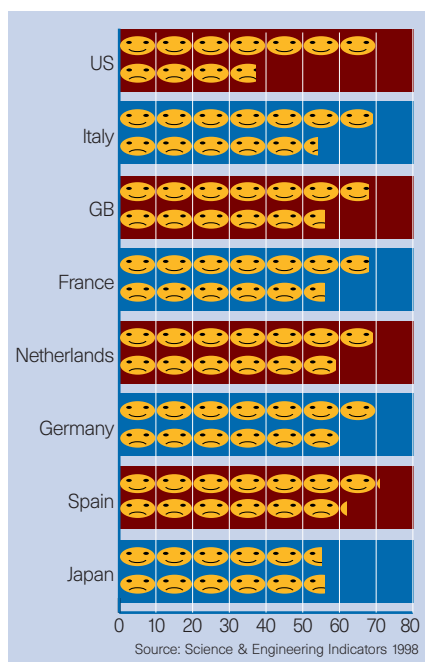
Miller, who has been conducting the survey for NSF since 1979, says the last finding reflects growing public familiarity with scientific ideas, partly as a result of strong science coverage in newspapers and on television. "It is a measure of how comfortable people feel with science," he says.

The survey also finds that Americans are vastly more positive about the overall impact of science and technology than those in Europe or Japan. Using a list of questions to generate an 'index of scientific promise' as a measure of positive attitudes, and an 'index of scientific reservation' to measure negative ones, the study indicates that Americans' positive impressions of science and technology have grown from an already high base.

The latest study gives the United States a 'scientific promise' index of 70 and a 'scientific reservation' index of 37, resulting in a 'confidence ratio' between the two indices of 1.89, up from 1.76 in 1992.

Numerous studies in European countries (and in Canada) have identified a similar belief in science's promise but far greater reservations, with resulting ratios of between 1.1 and 1.3 (see graph). A 1991 study in Japan indicated that people there share Europe's reservations but have much less faith in the promise of science and technology, with a resulting confidence ratio of less than one.

Miller attributes Americans' high confi-



Happy days: the ratio of those seeing high 'promise' to serious 'reservations' in science is higher in the United States than elsewhere.

dence in science to its successful record in assuring national security and better health-care, and in developing computer and telecommunications equipment. "This country has had a string of successes in science. For the average person, a lot of things have improved," he says. "And once you have a society that comes to celebrate its scientific achievements, that tends to perpetuate itself."

But the study found little improvement in the public's ability to grasp basic scientific

concepts. Only 11 per cent could explain what a molecule was, while 44 per cent knew that electrons were smaller than atoms. An index based such questions gave the United States a score of 55, similar to scores recorded in previous studies in the United Kingdom, Denmark and France, but higher than those of other European countries and Japan.

Although recent international comparisons have suggested that US schoolchildren are poor at science and mathematics, Miller believes that the reason US adults do quite well in his scientific literacy test is that 40 per cent of Americans go to college, and that nearly all of them are exposed to "at least a year" of science there. In Europe, as he points out, college students pursuing degrees other than science often get no exposure to it at all.

Miller adds that most of the public are unable to distinguish between science and technology, but says this is unsurprising. "If you asked a scientist whether the human genome project was 'science' or 'technology' you would get a good dinner-table discussion going, so you can't blame the public for being confused," he says.

Despite the strong support for science identified by the study, Miller worries that a sizeable proportion of the population still has little understanding of science.

The survey reveals that 45 per cent of respondents knew that the Earth goes round the Sun once a year. But any science educator who might lie awake at night fearing that most Americans retain the pre-Copernican view that the Sun goes round the Earth can sleep soundly: fewer than a quarter think this; the rest think the Earth goes round the Sun once a day.

Colin Macilwain

University says sorry for its racist past

[MUNICH] The University of Rostock has become the first east German university formally to apologize for removing academic titles on racial or political grounds during the Nazi era from 1933 to 1945. The university has said that the derecognition of academic qualifications was an "arbitrary measure" that was illegal and invalid.

The university has found in its archives 15 cases of derecognition, all involving Jewish male academics. Only one of them, Georg Cohn, still had a university post at the time. The others had fled Germany earlier, and so were automatically branded as political traitors by the Nazi regime.

The university's investigations also showed that derecognition of doctorates was

first proposed by German students, not by the Nazi government. In a letter sent to the Bavarian ministry of education in September 1933, the "German student body — Bavaria district" requested that the ministry strip of their doctorates "traitors who left the country". The proposal was eagerly pursued by ministries and universities all over Germany.

Angela Hartwig, head of the University of Rostock's archives, says that, when enquiries from *Nature* drew attention to the fact that the derecognitions had never been annulled, the university felt the need to compensate for the political injustice that had occurred during the Nazi era (see *Nature* 391, 112; 1998). **Quirin Schiermeier**

Indian meeting backed despite boycott threats over nuclear tests

[NEW DELHI] Organizers of the 10th International Congress on Immunology, due to be held in New Delhi in November, are shrugging off demands for a boycott from some members of the American Association of Immunologists (AAI) in protest at India's recent nuclear tests.

But several of those attending the meeting are still seeking to use the opportunity to register their disapproval of the tests and the international proliferation of nuclear weapons.

Eight foreign researchers have withdrawn their registration, but the organizers say they are relieved that a resolution was passed by the association last month expressing its support for the congress.

But they have been upset by a message circulated on the Internet by immediate past AAI president Charles Janeway of Yale University, urging the association not to support the travel of US scientists to India.

Janeway says he never intended to encourage a boycott of the conference, but adds that there is "a serious attempt, spearheaded by Indian scientists working in [the United States], and by myself, to attend the meeting and hold a special session on the Indian nuclear tests. We hope [to] maintain our scientific ties [with India] while protesting at the testing of nuclear weapons by India and Pakistan."

Abul Abbas, professor of pathology at Harvard Medical School, shares similar sentiments, but focuses on the dangers of worldwide nuclear proliferation. "It would be wonderful if scientists from India, Pakistan and all other nations could join to make a statement [at the congress] deploring the current situation," he says.

Some 1,430 overseas scientists have registered for the week-long congress, organized every three years under the aegis of the International Union of Immunological Societies.

The congress's president, Gursaran Prasad Talwar, an emeritus researcher at the International Centre for Genetic Engineering and Biotechnology in New Delhi, says he is pleased that many of the world's top immunologists plan to attend.

Although basic immunology will be the main focus, the New Delhi conference will for the first time devote three symposia and several workshops to what Talwar says is "beneficial immunology", including vaccines against cancer, allergies and autoimmune diseases such as arthritis. For the first time, the immunology of drug abuse and ocular disease will be discussed. Talwar says 108 'interactive' workshops have been designed to benefit young investigators. **K. S. Jayaraman**

Congress turns NIH budget into a political football

[WASHINGTON] An important Congressional subcommittee has proposed a 9.1 per cent budget increase for the National Institutes of Health (NIH). But it has done so at the expense of several high-profile social programmes, a move that has sharply divided Republicans and Democrats on the panel.

The virtually party-line vote two weeks ago by the Labor, Health and Human Services and Education subcommittee of the House Appropriations Committee suggests dramatic cuts to some social and education programmes. For instance, it would eliminate a \$1.1 billion programme of home heating assistance for the poor, and an \$871 million summer job training programme for youths.

The subcommittee also suggested halving the Goals 2000 education programme, a favourite of the Clinton administration. But it also seeks to boost the NIH budget by \$1.24 billion, to \$14.86 billion, and the Centers for Disease Control and Prevention would receive an unusually large 8.9 per cent increase. The bill continues to ban federal funding for human embryo research.

The bill scraped through the subcommittee by eight votes to seven. A corresponding Senate subcommittee must agree on its own bill, and the two versions must be reconciled and win approval from both legislative houses before becoming law before the new fiscal year begins on 1 October.

It is expected that some proposals will be rejected by the more politically moderate Senate subcommittee, which includes champions of the home heating programme and other affected social programmes.

Although Republicans have boosted NIH research — a politically popular position with both parties — Democrats nevertheless complain that it would take place on the backs of the poor.

President Clinton in a statement called the cuts "arbitrary" and "extreme". "The bill is out of step with our values and the wrong vision for America's future," he said. His comments were echoed by David Obey (Democrat, Wisconsin), the senior Democrat on both the subcommittee and the full Appropriations Committee.

Obey charged that Republicans on the panel "have decided to take the [NIH] funding out of the hides of the weakest and most vulnerable people in society". But he did not try to amend the bill to restore the social and education programme cuts and decrease the NIH boost.

Subcommittee Republicans, led by the chairman, John Porter (Republican, Illinois), insisted that cuts were being proposed not to finance medical research but because the



Porter: argues social policies deserve cuts.

social programmes did not deserve funding.

"This was not done capriciously, or to facilitate one kind of spending over another," says a spokesman for Porter. "These programmes did not merit continued funding."

For example, Porter argues that the home heating assistance programme begun during an energy crisis in 1980 is an anachronism now that energy prices have dropped, and that the summer job training programme has not made young people more employable.

But advocates for biomedical research praise the subcommittee's actions. "We are obviously pleased about the numbers for NIH," says David Moore, a lobbyist for the Association of American Medical Colleges.

Timothy Leshan, the director of public policy at the American Society for Cell Biology and a lobbyist for biomedical research funding, says that few scientists support cutting social programmes. But he adds: "there is political reality, and that's why they had to make the choices they did." The subcommittee must fund labour, education, social and health programmes from the same bill.

The money available for non-mandatory social and education programmes and medical research has remained virtually the same as last year because of a 1997 agreement that imposed caps on such 'discretionary' federal spending. Thus, to find an extra \$1.24 billion for NIH, and a \$208 million boost for the Centers for Disease Control and Prevention, the committee had to slash other programmes.

If the subcommittee's proposals prevail in their current form, the final version would risk a presidential veto in a Congressional election year. Rather than risk this, it seems likely that either the social spending will be restored to some degree in the Senate — and prevail in negotiations between the two bodies — or that the Republican leaders of Congress will increase the total money in the bill.

The latter could be achieved by tapping the growing budget surplus, or by breaching the spending caps in the 1997 agreement, which is allowed in narrow circumstances. "A lot of people believe there's going to have to be some more money put into that bill to resolve this issue," says Moore.

The full House Appropriations Committee is scheduled to meet next week to decide whether to endorse its subcommittee's recommendations. **Meredith Wadman**

Genome research set to take off in China

[BEIJING & SHANGHAI] China is launching several major initiatives in human genomics in Shanghai and Beijing, with budgets totalling over 250 million yuan (US\$30 million) over the next three years, and there are signs of even more investment to come.

New human genome centres are being established in both cities. Unusually for China, they are funded from a number of sources in both government and industry, and will draw researchers from several institutions. Both will focus on diseases that afflict large numbers of people in China.

The Shanghai centre is led by Zhu Chen of the Shanghai Second Medical University, and the centre in Beijing is under the leadership of Boqin Qiang, first vice-president of the Chinese Academy of Medical Sciences. Both have played an important role in China's human genome project, which until now has operated on a tiny fraction of the new budget.

The Shanghai Human Genome Center is located in a high-tech park in the city's new development zone of Pudong. It has an initial budget of 60 million yuan in the first year, with half coming from the Shanghai municipal government and Pudong.

The rest will come from the central government's Ministry of Science and Technology, the Chinese Academy of Sciences and various institutions that are participating in research at the centre, including Fudan University, Ruijin Hospital, Shanghai Medical University and the Second Military Medical University, as well as several institutes of the Chinese Academy of Sciences.

It is very unusual in China for scientists from so many different organizations affiliated with different parts of the government to work together under one roof. Some observers close to the project are sceptical about whether they will be able to collaborate successfully.

But Zhu Lilan, China's minister of science and technology, points out that the researchers have already been working together for two years under the human genome project, and says that Chen has the

ability to pull them all together.

The Beijing centre, which has a larger initial budget of almost 100 million yuan over three years, is also unusual in several ways. Two Beijing state-owned companies that invest in real estate and high technology will contribute over 50 million yuan to the centre, with the rest coming from the Beijing municipal government, the Chinese Academy of Sciences, the Chinese Academy of Medical Sciences, Beijing Medical University, Beijing Institute of Medicine and Pharmacology and the Ministry of Science and Technology.

Qiang says the central government hopes this unusual mix of funding from central and local governments and industry will provide a model for future funding of science in China, where the central government has limited funds but some parts of industry are booming.

The Shanghai centre will focus on identifying and cloning disease-related genes of particular significance to China, such as liver cancer. The centre's new building, which will open later this month, will expand within a year to a full complement of 35 to 40 researchers in four groups. These will work on expression profiles, genotyping for diseases, single-nucleotide polymorphisms and functional genomics.

Chen says the centre will also "coordinate" the various research efforts in human genomics in the Shanghai area. It will provide a high-quality service in genome analysis for research institutions and biotechnology and pharmaceutical companies at home and abroad, and serve as an "incubator" of genomics companies in China. It will make use of the rich human genetic resources of China, which has a fifth of the world's population and large numbers of patients with gene-related diseases.

Fudan University, which is participating in the research at Chen's centre, will maintain its own existing programme in human genome research, and the New Huangpu real-estate company in Shanghai will invest 100 million yuan in two laboratories at the university.



Looking to the future: Shanghai's Pudong zone where some of the research will take place.

Also, three US companies — PE Applied Biosystems, Axys Pharmaceutical and Siniwest Holdings — last year formed a joint venture in Shanghai called Shanghai GeneCore BioTechnologies to provide large-scale automated DNA sequencing and sequence analysis services for institutions in China.

This totally foreign-owned company has already won a large contract from the Ministry of Science and Technology to sequence human genes involved in liver cancer. Ming-Wei Wang, chairman and general manager of the venture, says the three companies will make an initial investment of about US\$10 million in three stages, in the form of cash, equipment and advanced technologies.

Other western companies also see Shanghai as an important centre for the development of the life sciences and the pharmaceutical industry. For example, Chen has received US\$1.8 million from SmithKline Beecham to support his research.

The Beijing centre will focus on different diseases from the Shanghai centre, in particular cardiovascular disease and diseases of the nervous system. According to Qiang, the centre will have a core of about 30 researchers working in three divisions — sequencing, genetic resources and bioinformatics — and 'satellite' laboratories in other institutions.

Qiang estimates that there are already about 30 laboratories in China working on the human genome project, with a total of about 500 researchers.

Several more health-related genomics projects are under consideration in China. The Ministry of Science and Technology will select several such projects for the National Programme for Key Basic Research Projects, otherwise known as the "scaling the heights" programme, which will provide about 50 million yuan per project.

David Swinbanks

Japan's Mars probe launched successfully

[TOKYO] Japan's Institute of Space and Astronautical Science last week launched PLANET-B, the nation's first Mars probe, from Kagoshima Space Centre on the southernmost tip of the island of Kyushu. The successful launch marks the beginning of Japanese efforts to join the United States and the former Soviet Union in the exploration of Mars.

The spacecraft will follow an elliptical orbit around Earth and the Moon until December, when it will begin its journey of

700 million km to Mars. PLANET-B, renamed Nozomi after the launch, is due to enter the martian atmosphere in October 1999.

The probe will spend two years observing interactions between the planet's atmosphere and ionosphere and the solar wind. The spacecraft carrying the probe also has a small aluminium plate bearing the names of more than 270,000 people following the institute's nationwide signature campaign (see *Nature* 391, 833; 1998).

Asako Saegusa

Japan okays test-tube baby gene tests

[TOKYO] Japan has taken its first step towards introducing preimplantation genetic diagnosis for human eggs that have been fertilized *in vitro*, with the approval last week by the Japan Society of Obstetrics and Gynaecology of guidelines covering such procedures.

The society's move ends a five-year standoff between members of the society and an anti-eugenics' patients' advocacy group over the approval of preimplantation genetic diagnosis. The procedure allows couples to ask for eggs used for *in vitro* fertilization (IVF) to be discarded if certain types of hereditary disorder are detected in the genes.

The society was last year forced to postpone the approval of guidelines for the genetic diagnosis of Duchenne's muscular dystrophy and fragile X syndrome in artificially fertilized eggs after opposition from the patients' advocacy group, which accused it of promoting discrimination against the disabled (see *Nature* 385, 763; 1997).

The new guidelines do not specify which diseases are to be tested for. The health ministry lists numerous hereditary disorders that can be detected by this method, but the guidelines say that each case will be assessed on its merits by a specialist committee of doctors and geneticists.

According to the guidelines, various factors will be assessed before the society gives approval for a genetic test, including the type of targeted disease, the genetic histories of the couple, and the facilities at institutions carrying out the test.

The society's existing ethical guidelines, which will now be modified, were drawn up in 1983 and restrict the use of IVF to treatments for infertility.

Kazuo Sato, president of the society, said in a statement that considerable discussion had been carried out with the public over the years, and that the guidelines, drawn up with special attention to the ethical issues surrounding preimplantation diagnosis, were intended to prevent misuse of the technique.

But members of the anti-eugenics patients' advocacy group, which represents more than 30 women's and disabled people's groups, complain that the society's decision has been rushed. One factor, they argue, was the recent furore over a doctor who successfully impregnated an infertile woman by IVF, using her husband's sperm and eggs donated by her sister. Such a practice is not allowed in Japan, as guidelines restrict IVF treatment to married couples, and prohibit the donation of ova and sperm from third parties.

Critics say the clinical application of reproductive techniques should be regulated by a more powerful body than the society. Members can only be expelled from the society if rules have been broken, and previous incidents have shown that the guidelines lack sufficient power of enforcement to prevent the abuse of techniques.

"The society made its decision rather abruptly, and discussions on the ethical and scientific issues seem to have been carried out in an ivory tower," complains Norio Fujiki, an emeritus professor of medical genetics at Fukui Medical School. "Discussions of bioethical issues related to medical techniques should not be confined to specialists — it should be debated on a national level."

Fujiki also says that preimplantation diagnosis should be regulated by law to prevent the risk of it being used for other purposes, such as selecting the sex of the child.

According to the Ministry of Health and Welfare, its council on health science plans to draw up new guidelines on reproductive techniques, including preimplantation diagnosis. But it says the potential social impact of the issue means it will need more time to consult the public.

Asako Saegusa

US think-tank queries cost of 'stockpile stewardship' programme

[WASHINGTON] The costs of US nuclear weapons research — including those of its new 'stockpile stewardship' programme — have been called into question by a report published last week. According to the report, the United States has spent \$5,500 billion on nuclear weapons since the start of the Manhattan project in 1940.

A 700-page study published by the Brookings Institution, Washington's pre-eminent liberal think-tank, says the cost of maintaining the US nuclear deterrent is now down to \$35 billion a year. Of this, \$4.5 billion is spent on 'stockpile stewardship', the programme designed to ensure the safety and reliability of US nuclear weapons.

Pointing out that annual expenditure on "activities now called stockpile stewardship" averaged \$3.6 billion (in 1996 dollars) between 1948 and 1991, the study, entitled *Atomic Audit*, questions why the programme's budget is higher than it was during the Cold War.

William Weida, professor of economics at Colorado College in Colorado Springs, and co-author of the report, argues that the programme has "huge amounts of money for construction projects which appear only to be intended to ensure continued employment at the national laboratories".

He is already working on a further study



that will "conservatively estimate" that stockpile stewardship "could be done for half or less" of the budget that the Department of Energy has allocated to it.

But Robin Staffin, deputy assistant secretary for nuclear weapons research at the Department of Energy, says the stewardship programme "is balanced, prudent and cost-effective".

Congressional appropriators in both houses have recently agreed to fund the programme next year at the requested level of \$4.5 billion. But Weida says the proposed budget for stockpile stewardship includes

large amounts of money for construction projects "whose uses are not yet specific, but which will evidently be dreamt up later" by the nuclear weapons laboratories.

Although the Clinton administration has said that the programme will cost \$45 billion over the next ten years, construction costs of planned facilities tail off after the year 2003, leaving hundreds of millions of dollars for additional facilities.

Staffin argues that it is normal for the Department of Energy to reserve money in its budget projections for undefined construction projects. He adds that the department is exploring the need for various such facilities, although it has not yet given its backing to any.

One proposal is a successor to the Dual-Axis Radiographic Hydrodynamic Test facility at the Los Alamos Laboratory in New Mexico. Another is a pulsed-power fusion device at the Sandia Laboratory, planned for possible construction on the mothballed nuclear weapons test site in Nevada.

More than half of the money Brookings says has been spent on nuclear weapons since 1940 — \$3,200 billion — has gone on deploying weapons systems. Research, development and production has cost \$410 billion; the clean-up of nuclear facilities has consumed just \$45 billion.

Colin Macilwain

Public 'must have more say on NIH spend'

[WASHINGTON] An expert panel told the US National Institutes of Health (NIH) this week that it should give the public a bigger voice in deciding how to spend its \$14 billion annual budget. The panel's suggestions include setting up public liaison offices in all 21 of its institutes, as well as in the office of the NIH director.

The panel also calls for more diverse public representation on the director's advisory committee, and a stronger role for the committee in setting research priorities. And it recommends the creation of a Director's Council of Public Representatives to bring public opinion to the director.

In its report published on Wednesday (8 July), a committee of the Institute of Medicine (IOM) calls on the NIH to introduce broad changes to the way it communicates with the public about research decisions. The agency "must revamp its approach to public input and outreach—at every level—without delay," writes the committee chairman, Leon Rosenberg, in the report. Rosenberg is a professor in the department of molecular biology at Princeton University.

"By creating formal links to the general public, NIH can ensure that all have a voice in what gets funded, and that more people understand how such decisions get made," he says. Such changes "would underscore that openness is as important to the process as expertise and objectivity".

The 120-page report, *Scientific Opportunities and Public Needs: Improving Priority Setting and Public Input at the National Institutes of Health*, was prepared by a 19-member committee that heard from researchers, NIH staff, disease advocacy groups and members of the public over five months.

It was mandated by Congress last year in a move reflecting the view of some members of Congress that the study would provide a needed line of defence for NIH against Congress's tendency to legislate on the details of research spending. They reasoned that the report could recommend changes in the way that NIH sets its priorities that would ultimately give less cause to Congress to meddle.

Disease advocacy groups have convinced other members of Congress that their causes need funding that is congressionally mandated because NIH's decision-making process is too closed.

The IOM report agrees that the NIH should consult a broader range of opinion: "NIH, especially the Office of the Director, does not have adequate channels through which the public can provide broad input into the NIH priority-setting process, or through which NIH can respond clearly to the public on issues of mutual concern."

The report "generally" endorses the criteria by which NIH allocates research dollars as

"reasonable and appropriate". But it recommends that the agency should develop data comparing the costs of specific diseases against the resources devoted to them.

NIH officials have told Congress that such data is highly complex and can be misleading, partly because basic research in one field can yield unexpected applications in others. But the panel says that NIH should come up with reliable figures anyway.

"We believe it can't be brushed aside," says Rosenberg. "If [the numbers] are done systematically and seriously, confidence will grow that this is a legitimate attempt."

The report recommends — as did an IOM study in 1984 — that the NIH director should have greater authority to control spending priorities throughout the agency, in part by requiring each institute to submit strategic plans annually. **Meredith Wadman**

Panel casts doubt on helium sell-off plan

[WASHINGTON] A plan to privatize a vast reserve of helium that the US government holds in an old natural gas dome under Texas may be impossible to implement, according to the chair of a National Research Council (NRC) panel. The panel is about to investigate the impact of the proposal on scientists and other users of the gas.

In October 1996, President Bill Clinton signed a law mandating the sale of the helium reserve, which holds 30 billion cubic feet of the gas at Amarillo, Texas. The legislation instructs the government to sell it between 2005 and 2015 at a price which, most analysts believe, the market cannot bear. It also requires the government to keep 600 million cubic feet, excluding the possible transfer of the whole reserve to a private company.

"This is going to be an interesting sale, because the whole thing isn't for sale," said Ray Beebe, the Tucson, Arizona-based consultant who co-chairs the NRC panel, at its first meeting on Monday (6 July). "Is the legislation a series of compromises, to the point where it might not be workable?" he asked.

The study was requested in the legislation as a concession to groups such as the American Physical Society, which believes that the government should keep the reserve for future

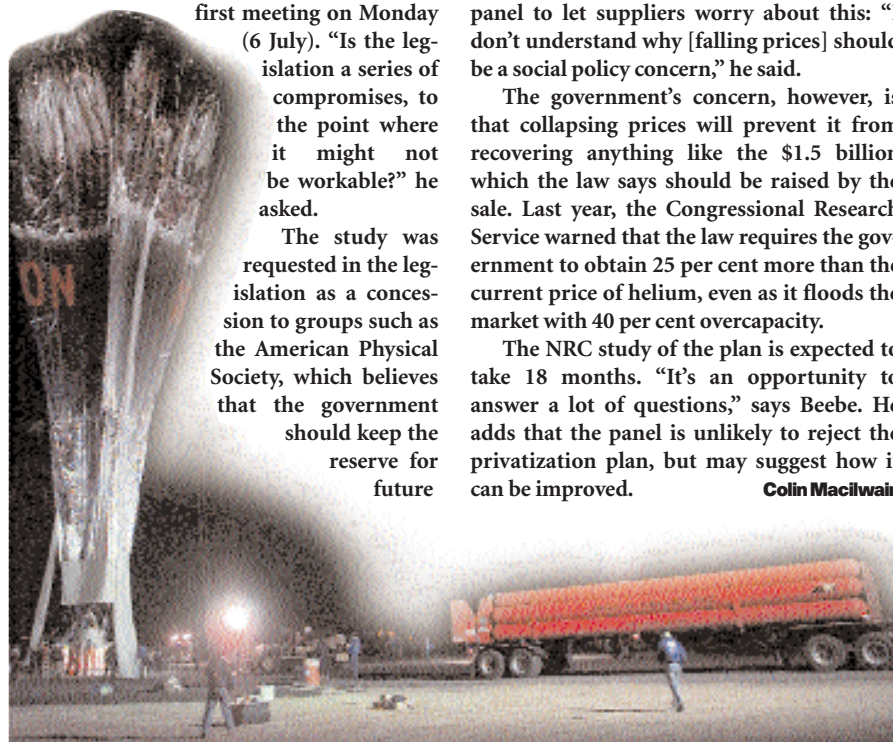
use. Physicists say they want to protect the reserve because of the extreme difficulty of extracting the gas from anywhere other than a few, non-renewable natural gas fields. "The physics community is aware of all sorts of technologies that could depend on helium," says Allen Goldman, a panel member and physicist at the University of Minnesota.

The United States produces three-quarters of the 4.5 billion cubic feet of helium that the world will consume this year. Available global reserves are estimated at 200 billion cubic feet. Use is growing rapidly, but the price of helium has been stable in recent years, at around \$35 per 1,000 cubic feet.

Apart from concern about consuming a non-renewable resource, critics worry that privatization of the reserve will flood the market, slashing prices and leading commercial suppliers to stop recovering helium from natural gas. But Tim Brennan, an economist at the University of Maryland, advised the panel to let suppliers worry about this: "I don't understand why [falling prices] should be a social policy concern," he said.

The government's concern, however, is that collapsing prices will prevent it from recovering anything like the \$1.5 billion which the law says should be raised by the sale. Last year, the Congressional Research Service warned that the law requires the government to obtain 25 per cent more than the current price of helium, even as it floods the market with 40 per cent overcapacity.

The NRC study of the plan is expected to take 18 months. "It's an opportunity to answer a lot of questions," says Beebe. He adds that the panel is unlikely to reject the privatization plan, but may suggest how it can be improved. **Colin Macilwain**



Balloon goes up: helium use is on the rise, but US privatization plan could flood the market.

AP/ERIC DRAPER

CITES chief removed in scandal over trade in banned species

[LONDON] Izgrev Topkov, secretary-general of the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES), has been removed from his post after six years at the helm of one of the United Nations' most controversial secretariats.

Two members of the CITES secretariat in Geneva have also been dismissed for their role in awarding permits to organizations that wanted to trade in plants and animals on the CITES list of banned species.

The dismissals have been authorized, following two investigations, by Klaus Töpfer, the executive director of the United Nations Environment Programme (UNEP), which is responsible for the CITES secretariat.

A UNEP spokesman confirmed that Topkov has "been moved", and that two of his staff have been dismissed "for improper behaviour in handing out permits". But Topkov's eventual fate is uncertain; he may take up another post within the UN.

Topkov was personally not implicated in any wrong-doing, according to UNEP sources. But his failure to prevent the permits being given out for trading in banned species meant that he had to take ultimate responsibility. "If you have responsibility to manage something, and you are unable to exercise that authority properly, you have to go," sources say.

Topkov's deputy, Jim Armstrong, has taken over temporarily as secretary-general of CITES. Topkov's responsibilities will eventually be looked after by a senior UNEP official until a more permanent replacement can be found.

The unauthorized granting of permits is believed to have taken place over at least two years. Two separate investigations — one by a CITES committee of country representatives, and another by the UN — are understood to have reached the same conclusions.

Topkov is a former Bulgarian diplomat, and became secretary-general of CITES six years ago. He was previously a member of the intergovernmental negotiating committee responsible for the UN climate convention.

The investigation into the CITES secretariat was completed last December, and Topkov's position has been looking increasingly precarious ever since. But a decision about his fate could be taken at the time only by the then executive director of UNEP, Elizabeth Dowdeswell, whose own contract was coming to an end.

CITES is one of the oldest UN conventions, having come into force in 1975. Its 143 member states agreed to ban commercial international trade in endangered species listed by CITES.

Ehsan Masood

Small-scale space projects 'threaten collaboration'

[MUNICH] A transatlantic panel of space scientists has warned that attempts to boost international collaboration among space agencies are being challenged by the shift to 'smaller, faster and cheaper' missions. Collaboration is seen as increasingly necessary because of shrinking budgets.

This is one of the conclusions of a report by the European Space Science Committee, a body associated with the European Science Foundation and the US National Academy of Science's Space Studies Board. The report was presented last week to the US space agency NASA, the European Space Agency (ESA) and national space agencies.

International collaboration in space missions saves money and promotes good front-line science, says the report. But NASA's new philosophy of launching fewer big missions and more 'smaller, faster, cheaper' missions — now being followed to some extent by ESA — discourages collaboration because of the rapid turn-around times of small missions and the reduced financial incentive to join forces.

The report summarizes studies of 142 cooperative projects and missions between the United States and individual European countries or ESA that have taken place in the last 30 years.

The report says that most international missions, such as the Hubble Space Telescope and the SOHO solar mission, have been scientifically very successful. Hubble and SOHO were collaborations between NASA and ESA. Like the Germany/UK/US X-ray astronomy mission ROSAT, Hubble and SOHO were driven by clearly defined scientific goals, and backed by strong scientific communities on both sides of the Atlantic.

But the committee points out that a significant number of projects were not fully successful in meeting their scientific goals. It identifies eight key elements that could be used to assess whether future international missions are likely to be successful.

These include scientific support through peer review, clearly defined scientific goals and a strong scientific community, shared objectives between all partners, appropriate procedures for data handling and distribution, and a "sense of partnership".

The report highlights the ESA-NASA gamma-ray astronomy mission INTEGRAL as an example of a shaky collaboration. INTEGRAL was almost abandoned in an advanced development phase because of an unexpected decision by the US Congress not to fund one of its the mission's main instruments.

The strong European gamma-ray astronomy scientific community managed to patch things up, and INTEGRAL is now due to be launched in 2001.

The error in setting up the international collaboration, says the report, was that a scientific community of equivalent strength did not exist in the United States. That made the mission vulnerable to political and budgetary attack there.

In contrast, the major problems that almost inevitably occur in complex missions were able to be handled with ROSAT and SOHO because of their strong and well-coordinated scientific teams, says the report. The ROSAT hardware had to be modified at the last minute when the mission's Shuttle launch was abandoned because of the Challenger disaster, and the satellite had to be redesigned to interface with an expendable launch vehicle. But this caused little delay to the project's schedule.

Formal procedures, such as the use of standardized agreements, would help to avoid problems of misinterpretation of ad hoc agreements, and of partners pulling out of missions in their development stages, says the report.

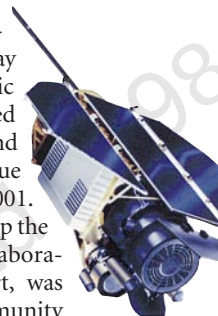
Clear advance agreements on data handling are particularly important for Earth observation missions, where security and commercial interests have sometimes hindered data exchange.

A sense of partnership is a "recipe for success", says the report, making it important that the media and scientific reports should acknowledge all the partners concerned. European scientists have frequently been angered by the neglect of their contributions to Hubble observations. ESA contributed 15 per cent of Hubble's development and construction costs, and wins 20 per cent of observation time through peer-reviewed competition.

John Culhane, director of the Mullard Space Science Laboratory in London and chairman of the European Space Science Committee, says he hopes the report will encourage the agencies to develop more systematic approaches to international collaboration.

"Science can prosper from collaboration independent of the issue of saving money because it brings more creative thinking to bear on missions," he says.

Successful teamwork: ROSAT redesign caused little delay.



ROSAT

Future of Gabon lab hangs in the balance

Declan Butler

The choice of a new director for the International Medical Research Centre in Gabon is crucial for the future of an institution which has been criticized as lacking relevance and falling far short of its potential.

[PARIS] Well-equipped and properly staffed research centres are rare in Africa. It is therefore not surprising that tropical-disease scientists have their eyes firmly fixed on the outcome of a stormy period at the International Medical Research Centre (CIRMF) in the former French colony of Gabon, one of Africa's biggest French research investments.

In particular, the imminent choice of a new director will be crucial to the CIRMF's future. The institute is widely considered to have the potential to rank as an international centre of excellence in infectious diseases, but its ambitions have been hampered by chronic administrative and social strife.

The centre has lacked a director since last July, when it declined to renew the mandate of Alain George, a French scientist who quit the country after disputes with local authorities. According to several scientists familiar with the centre, tension between local management and outside researchers is rife. They claim this is handicapping efforts to create a strong research programme by, for example, prompting the departure of several leading scientists. The climate also appears to deter people; several top scientists, for example, have turned down the directorship.

Colonial past

The centre's scientific board recently selected Michel Blanchot, a pharmacologist at the military hospital of Val-De-Grâce near Paris, as candidate for the post, and rejected the only other applicant, Robin Ryder, an experienced virologist at Yale University in the United States. The executive board, which consists largely of state officials, is expected to take a final decision during the summer.

Blanchot last week declined to comment on his selection, pointing out that this may yet be rejected by the executive board. He is widely considered to have broad administrative experience and an in-depth knowledge of Africa, although he is no longer active in research.

"A strong and fresh outlook is needed," says one member of the CIRMF scientific board. He argues that the new director will have to combine "meticulous attention to world-class science" with the sensitivity needed to work in an Africa that is strongly



Under scrutiny: the centre is well equipped but its operation has long been controversial.

resentful of past colonial attitudes.

In addition to its research activities, CIRMF — which has around 170 staff — has from the outset been subject to strong political influence. It was founded in 1974 at the request of Gabon's president, El Hadj Omar Bongo, largely through the financing of the giant French oil company Elf Aquitaine.

The company, whose financial and political muscle helped bring Bongo to power, and whose revenues make up the bulk of the country's wealth, agreed to create the centre as part of a deal giving France access to Gabon's oil fields.

Luxuriously furnished, with a well-equipped primate centre and a generous fleet of Land Rovers, the centre — which has an annual budget of FF40 million (US\$6.5 million) — is based in Franceville, the president's home town, 1,000 km from the capital and accessible only by light aircraft.

Several scientists claim that there is widespread local nepotism and abuse of power. According to one researcher, "people are vying for power and influence at the centre, and strategy is not based on research excellence, but on political and family connections".

Social tension at CIRMF has been exacerbated by the devaluation of the African French franc. This has substantially reduced the salaries of African staff which were already low compared with those of foreign researchers, who are paid in hard currency.

Although CIRMF has a respectable list of publications in leading journals, particularly in primatology and virology, there is a broad consensus among many tropical-disease researchers that its way of operating prevents it from becoming a leading African and

international research centre.

"The centre has not reached its potential, which is far greater than its current output," agrees Joseph McCormick, a leading virus 'hunter' formerly with the Centers for Disease Control in Atlanta and now head of an epidemiology programme at the Institut Pasteur in Paris.

McCormick, who is also a member of the CIRMF scientific board, says the centre is uniquely placed geographically to exploit the "huge" scientific possibilities offered in Africa, the cradle of infectious diseases such as Ebola and HIV. "[The region] is where we are going to understand the origins and ecology of viruses and other disease agents," he says.

Local difficulties

One member of the CIRMF science board, Tom Nchinda, an African who is a former head of the World Health Organization's capacity-building division, also argues that the training of local scientists must be given much greater priority. This is the key to sustainable development of an African science base, he argues, adding that "CIRMF is not fulfilling its role as an institutional base compared with other centres in Africa".

Criticisms of the centre's organization are vehemently dismissed by Joseph Lansoud Soukate, its interim director. "The centre is doing very well, we do not have problems," he says. "We are not in the process of collapsing as some people would have you believe," he says, referring to president Bongo's support for the centre. Soukate declines to comment on the substance of the criticisms.

Henri Mathieu, chairman of the centre's scientific board and president of the Hôpital Robert Debre in Paris, admits that criticisms of the centre are "not entirely without foundation", but he cautions that these must be seen in the context of the difficulties of working in Africa.

Mathieu also defends the centre's focus on long-term fundamental research rather than on applied research of greater immediate relevance to local public health needs. He argues that studies on the genetic polymorphism of infectious diseases, for example, will ultimately have significant implications for public health.

But he concedes that the centre also needs to make an "extraordinary effort" to improve the local application of its research if it is to persuade the population and politicians of its usefulness.

According to one official at the World Health Organization, many politicians outside the presidential entourage are increasingly critical of what they claim is the centre's lack of relevance to Gabon's public health needs and its relative isolation from the rest of the country's health and university infrastructure. Some are reported to be keen to see oil revenues diverted elsewhere. □

No plan to privatize Russia's universities, says prime minister

[MOSCOW] Sergei Kirienko, the Russian prime minister, has told the heads of Russian universities that, contrary to earlier rumours, the government has no plans to privatize the country's universities. But he has emphasized that his top priority for higher education is the quality — and not the quantity — of the graduates they produce.

Addressing the congress of the Russian university rectors union, Kirienko also said that, if a university wished to produce more graduates than were needed by the state, it would have to find its own resources to do so. He promised that the government planned to pay all the money it owed to universities, and that for the next two months it will cover all their service costs.

Holland hosts school for study of climate change

[MUNICH] A new International School for Cooperation on Oceanic, Atmospheric and Climate Change Studies opened last week in Utrecht, the Netherlands. The school, which is a joint German–Dutch initiative, will develop methods for measuring greenhouse gases, and develop and verify climate models through summer schools, specialist courses and workshops.

Information and computer capacity will be supplied by the German Climate Computer Centre in Hamburg and several computer centres in the Netherlands. The school has been funded by the German and Dutch ministries of research to the tune of DM4.4 million (US\$2.4 million) a year for up to nine years. It will be run by scientists from several climate research institutes in the two countries.

Levine takes top job at Rockefeller University

[WASHINGTON] Rockefeller University in New York has chosen as its new president Arnold J. Levine, currently professor of molecular biology at Princeton University, who is perhaps best known as a co-discoverer of the p53 tumour-suppressor protein. He will succeed Torsten Wiesel, who retires later this year, having taken over the reins from David Baltimore in 1992.

Levine said last week that he was “privileged” to join Rockefeller University “during such a challenging and exciting time for this great institution”. Wiesel described Levine — who headed a major review of the AIDS programme at the US National Institutes of Health (see *Nature* 380, 190; 1996) — as a “superb investigator and

academic leader whose knowledge and interests extend well beyond his own area of expertise”.

Joliot to chair French science ethics panel

[PARIS] Pierre Joliot, a professor at the Collège de France in Paris, has been appointed chairman of COMETS, the consultative ethics committee created in 1984 by the French Centre National de la Recherche Scientifique. He succeeds the historian, Hélène Ahrweiler.

Joliot's standing in French science — he is the son of Frédéric Joliot, Nobel prizewinner for chemistry in 1935 — should “give him a strong capacity to resist undue pressures”, says one former member of COMETS. He points out that the committee covers such politically sensitive issues as the internal operation of the research organizations (see *Nature* 388, 110; 1997), scientific misconduct and revisionist historians.

The price wasn't right for Einstein's letters

[LONDON] A much-anticipated and widely publicized New York auction of Albert Einstein's memorabilia, including love letters to the alleged Soviet spy Margarita Konenkova, fizzled out last week when the lot failed to attract its expected \$250,000 price. The highest bid was \$180,000.

The Einstein items included nine letters, an engraved gold wrist-watch and photographs of the physicist. The lack of interest is in contrast to an auction in 1996 when a collection of Einstein's letters to his first wife, Mileva Maric, fetched \$900,000. The same year, a manuscript of his Special Theory of Relativity sold for \$5 million.

Supercomputer service on way for UK academics

[LONDON] The British government is to set up a £26 million (US\$43 million) supercomputing service for academic researchers. The service will provide access to a supercomputer capable of performing 700 billion calculations per second. This makes it the most powerful computer in Europe for academic research. It will be upgraded to twice its present power in three years.

The computer will be built with private finance, with users paying according to the amount they use the service. The design will be based on the latest T3E-1200E supercomputer manufactured by the company Silicon Graphics, one of the service's three-member consortium. Other members are Computer Services Corporation, which will manage the service, and the University of Manchester, where the service will be based.

US gene sequencers step up the pace

[WASHINGTON] US human genome sequencers will aim to sequence 117 million bases — nearly twice what they have achieved to date — in the 12 months from 1 July, funded by government grants announced last week.

The National Human Genome Research Institute (NHGRI) said it will spend \$60.5 million at seven sequencing centres.

Dropped from funding was The Institute for Genomic Research in Rockville, Maryland, which announced in May that it is teaming up with Perkin-Elmer in a privately funded project to sequence the genome (see *Nature* 393, 101; 1998).

Mark Guyer, assistant director for scientific co-ordination at NHGRI, says that the funding decisions leaves the institute “confident that we are on target for producing the kind of [completed] sequence that we had targeted”.

Dolly researchers win grant to continue work...

[LONDON] The Roslin Institute in Edinburgh and the company PPL Therapeutics have been awarded £600,000 (US\$989,000) by the British government to improve on the technology that was used to clone Dolly the sheep (see *Nature* 385, 810–813; 1997). The grant, which was awarded under the Department of Trade and Industry's Cell Engineering Link Programme, will support a three-year project aimed at improving the precision of genetic modification techniques in nuclear transfer.

Genetic modification will allow the institute and PPL Therapeutics to clone animals with specific genetic characteristics. Last year there was considerable publicity around the fact that an earlier grant from the agriculture ministry to the institute came to an end just as the birth of Dolly was stirring up worldwide publicity.

...as Japanese announce cloned calf twins



[TOKYO] Japanese scientists said this week that they had produced the world's first cloned calves (left) using

techniques similar to that used last year in Britain to produce Dolly, the cloned sheep.

The joint research team at Ishikawa Prefectural Livestock Research Centre and Kinki University announced the birth of twin calves produced by cloning somatic cells taken from the uterine lining of an adult cow.



Museum research comes off list of endangered species

Natural history museums are shaking off their dusty image in a bid to show relevance to contemporary concerns. Central to a revival in their research fortunes is a unique contribution to our understanding of life's complexity.

Last month, the American Museum of Natural History in New York, perhaps the international flagship of museum research, opened a public Hall of Biodiversity (above), the driving force behind which is a new multidisciplinary Center for Biodiversity and Conservation.

The centre symbolizes the faith that the directors of the world's major natural history museums, whose commitment to whole organisms in an era of cell and molecular biology means they are often perceived as 'dens of antiquity', are placing in biodiversity

to revive their fortunes. It also reflects the growing awareness, after decades of neglect by public funders, of the relevance of museums to underpinning global, regional and national efforts to conserve biological diversity. The museums' vast repositories of biological specimens and associated expertise make them ideally placed for this role.

"It's hard to compete with a place like this in biodiversity," says Michael Novacek, provost of the New York museum. He says that decades of falling interest in whole-organism biology within the universities

This Briefing has been written by Declan Butler in Paris, with additional reporting by Henry Gee in London and Colin Macilwain in Washington.

have left museums in a position to excel as centres of knowledge on biodiversity.

So far, the political prominence given to biodiversity has failed to translate into substantial financial support. "There is a great deal of talk in the biodiversity movement of the need for more research, but it has not led to large increases in funding and staff," says Sir Ghilleen Prance, director of the Royal Botanic Gardens at Kew in London.

Nevertheless, many museum directors are optimistic that biodiversity is the best bet in a long time for making a return to political favour. "I am convinced that the situation is about to change," says Pieter Baas, director of the Rijksherbarium in Leiden, the Netherlands, one of the world's largest herbariums.

Baas says there are already signs of an upturn in funds for biodiversity research in several countries. "There is always a certain time lag before politicians agree to a major investment in research," he says. "This is all the more so when we are talking about disciplines that have been around for centuries."

Such optimism is based, for example, on the unequivocal recognition of the value of museum research disciplines at the fourth meeting of countries which have signed up to the United Nations Biodiversity Convention, held in Bratislava in May of this year.

The 175 signatories adopted a Global Taxonomy Initiative which acknowledges that there is a 'taxonomic impediment' to the sound management and conservation of biodiversity. Removing this, they agreed, is a crucial, rate-determining step in the implementation of the convention's objectives.

The statement also affirms the need for more taxonomic experts, and for a stronger infrastructure to discover and understand the multiplicity of relationships among species of plants and animals.

Many scientists consider that such statements provide a clear political mandate for funding a series of integrated research programmes in biodiversity, for renovating their research collections — many of which are in a dire state of decay (see over) — and for injecting much-needed young blood into collections-based research.

Museum collections are to biodiversity research what genome databases and cell libraries are to genomics. The collections held in the world's 6,500 or so natural history museums — which total about 3 billion specimens, not counting microorganisms — represent the work of thousands of individuals carried out over centuries.

Part of their value lies in the fact that, before a new species can be described, it must

be compared with the 'type' specimens held in museums, and named according to prescribed taxonomic rules. The need for a rational method of organizing biodiversity requires the discipline of systematics, which arranges taxa — such as species and genera — into classifications (that may or may not reflect their phylogenetic relationships).

"Collections are absolutely essential to biodiversity research," says Prance. Creating

a complete inventory of life on Earth is impossible without reference to museum specimens, while the associated taxonomy and systematics is needed to make rational decisions about conservation.

Imagine, for example, that an oceanic state wishes to develop conservation policies for an island having ten endemic bird species. If systematic studies show that nine of the species are close relatives, but that the

tenth represents a distinct lineage, the most cost-effective policy might be to focus conservation measures on protecting the tenth species, as its loss would be more significant than that of one of the nine related species.

Mexico is already applying such logic to the state of Veracruz, whose biodiversity has been best described. It is trying to build a picture of the abundance and distribution of the state's biota, and its endemic species, using information derived from museums worldwide on 137,500 specimens of 7,300 species — including butterflies, birds, mammals, reptiles, amphibians, plants, fungi, earthworms and dragonflies. This baseline data is being used to plan ecological management and conservation strategies.

Some are already seeking to extend this to a global scale. For example, Systematics Agenda 2000 is a proposed US\$3 billion initiative that aims to understand the 'big picture' of evolutionary history, by discovering, describing and classifying every species on Earth, so completing the research tasks begun by Darwin and Linnaeus.

The proponents of Systematics Agenda 2000 argue that, although 1.75 million species have been described, the total number is thought to be somewhere between 6 million and 100 million (according to a Global Biodiversity Assessment published by the United Nations in 1994). And, with just 15–20,000 new species being described annually, a rapid acceleration is needed.

"We are talking about exploring the entire biosphere of a whole planet, and ideally within a few decades, as much of that diversity is going to disappear," says Quentin Wheeler, a researcher at Cornell University and a former president of the Association of Systematics Collections. "Imagine you told people it was Mars, and that there were 100 million species there and that as many as a quarter might be gone in 30 years — there would be a huge outcry to launch missions and learn about it before it was gone."

The programme was launched by the American Society of Plant Taxonomists, the Society of Systematic Biologists, and the Willi Hennig Society. According to many observers, it lost some of its initial momentum when museums and other participating institutions became preoccupied with their own financial difficulties.

But it has recently regained prominence following its absorption into a broad international umbrella biodiversity organization, Diversitas, set up under the auspices of the United Nations Educational, Scientific and Cultural Organization, and the International Council for Scientific Unions, and sup-



Blackmore: focus on cooperation under Diversitas.

Funding for renovation is the top priority



Flying fish: successful refurbishment at France's Muséum National d'Histoire Naturelle.

US researchers visiting the Muséum National d'Histoire Naturelle in Paris are surprised to learn that they can leave transformers for their binocular microscopes and other electrical devices in their hotel rooms. The museum's electrical system is antiquated — like much of the rest of the museum — and still operates at 110 volts, whereas the rest of Europe switched to 240 volts decades ago.

The dire state of the vast museum is an extreme example of a problem facing many of the world's natural history museums. Their collections are deteriorating, and there is an urgent need for expensive and well organized storage and research conditions to keep specimens free from insects, dust and extremes of temperature and humidity.

An example of what renovation can achieve comes from the Smithsonian Institution in Washington, DC, which has transferred 20 million specimens from the museum's cramped quarters to a purpose-built 50,000-square-metre storage and research centre in Suitland, just south of the capital. This is a far cry from the Paris museum, where collections are often piled in boxes in corridors, and where a fire in 1995 came within metres of one of the world's richest herbariums, a tinderbox of 8 million specimens (see *Nature* 385, 378; 1997).

A recent report by France's interministerial committee for national public works warns that the museum will face a crisis unless the government launches a massive refurbishment programme. "If no new investments are made, tomorrow it will be the laboratories that face closure," it says.

The Paris museum's collections are

rivalled only by those of the Smithsonian Institution and London's Natural History Museum. They include 2 million fossils, 150 million insects, and 1.5 million vertebrates, including 1 million fish, 200,000 amphibians and reptiles, and 272,000 birds and mammals — not to mention a myriad of geological specimens including seven kilometres of 'carrots' from ocean drilling.

Henry de Lumley, the museum's director, says his "number one priority" is to renovate the collections, which currently receive just FF2 million (US\$330,000) a year. The report will provide him with ammunition for his proposal to the state's major public works funds for a FF2 billion, 15-year renovation programme, similar to the FF500 million renovation of the museum's Great Hall of Evolution (see *Nature* 362, 280; 1993).

Refurbishment of collections is also a burning issue at the Natural History Museum in London, even though its finances are in much better shape than those of the Paris museum, following a decade of harsh austerity measures (see page 119). "We are really concerned about the state of our collections," says Steve Blackmore, keeper of botany. "Establishing the need for that expenditure is something that we need to continue pressing our case for."

The London museum is able to pay for a planned £300,000 (US\$500,000) rehousing of its 6 million herbarium specimens from its own resources. But an urgently required 15-year renovation of its other collections housed in unsuitable Victorian buildings — estimated to cost £60–80 million — is well beyond the museum's means, and will require a special fundraising effort.

ported by many museums and research institutions worldwide.

"Many people are focusing their support on Diversitas as an all-embracing initiative," says Steve Blackmore, keeper of botany at the Natural History Museum in London. Prance, who chairs Diversitas, says that what sets it apart from other conservation initiatives is that it exists only to "promote and co-

ordinate research on biodiversity". Diversitas already plays an important role in co-ordinating the activities of museums and other research institutions internationally. It has set up working groups to assess the state of knowledge, identify gaps, and reach international consensus on priorities for taxonomic research — for example on groups of organisms that should be targeted.

One of those who endorses the need for a coordinated international approach is Dennis O'Connor, provost of the Smithsonian Institution in Washington, DC. "It is going to be very important for the organizations involved in systematic biology to talk about what their priorities are, and how they all can participate," he says. "We won't get there by conducting business as usual."

Core missions under growing pressure

The demands of biodiversity research are leading to a new assertiveness among museums that they should be respected as places of scholarship uniquely well placed to pursue a holistic, multidisciplinary approach to fundamental research on the Earth and its biosphere.

Broadly, natural history museums have three core missions: conservation of collections, public understanding, and research. Museum research is distinguished from that carried out in universities by the fact that it is built around the collections.

But this specialized role has long been vulnerable to two sets of pressures. One is the need to seek research funds from outside agencies whose goals are often very different from those of museums. The other is the desire of funders — and often of museum management, acting at their behest — to devote resources to meet the public's demand for increasingly costly, blockbuster, theme-park style exhibitions.

"It must be recognized that building attendance and developing strong programmes in collections and research are not mutually exclusive," says Peter Crane, director of the Field Museum in Chicago. "The great natural history museums will not survive if collections and their associated research become peripheral to their future."

The Muséum National d'Histoire Naturelle in Paris has recently been criticized, in a report by an interministerial committee, for failing to assert its core missions sufficiently. These are to create an inventory of biodiversity using the techniques of systematics, genetics and molecular biology; to study the evolution of man and the biosphere; and to promote understanding of the environment. The report complains that researchers at the museum employed or funded by external agencies, such as the Centre National de la Recherche Scientifique, are assessed not on their contribution to the museum's missions, but solely on the basis of their publication record.

This narrow assessment has led researchers to pursue work that is often distant from the museum's core missions, says the report. It has also resulted in a neglect of collections-based research and curatorship. The report calls for new statutes that define

clearly the allegiances of researchers funded by external agencies, and ensure that external bodies take into account the specialized role of museum researchers, and their contribution to the museum's missions.

Henry de Lumley, the museum's director, contests this analysis, arguing that the museum's modern research laboratories fulfil an important logistical role within the institution. But he agrees that there has been an overemphasis on medical research in some labs, and that funding agencies have in the past placed excessive emphasis on molecular biology projects, for example, at the expense of systematics and whole-organism biology.

Quentin Wheeler, from Cornell University, New York, a former president of the Association of Systematic Biologists, says he accepts that molecular biology and other modern disciplines are essential to a museum's activities. But he says that the overriding need of cash-strapped museums to find funds has often resulted in the growth within them of research areas that have little relevance to evolution or systematics.

Rather than asserting their core missions, Wheeler says museums have tended "to follow the orientations of outside grant agencies and universities; like moths on a pheromone trail, they see the dollar signs". He argues that many museum researchers are doing work that could be carried out at any university, such as medically-oriented research on natural products.

Even within core museum research areas such as systematics, tensions often run high between traditional and modern sciences. "Our local molecular lab has damn little to do with any of the museum systematists; furthermore, they compete for space, fellowships, and funding," protests Victor Springer, Axelrod Curator of Fishes at the Smithsonian Institution's National Museum of Natural History in Washington, DC.

Much of such antagonism can be traced to resentment by traditional systematists at the high costs of such facilities as DNA sequencing laboratories, when they have few funds to maintain their collections, argues Edward Theriot, director of the Texas Memorial Museum in Austin.

Theriot is both a museum administrator — "with a responsibility to seek opportunities for support" — and president of the Association of Systematics Collections. He admits that museums face a difficult task in balancing the need to be involved in cutting-edge research, such as molecular systematics, with their duty to carry out research based around their collections.

But, in opposing molecular biologists, traditional systematists are "attacking the wrong enemy", argues Theriot. He says the overriding need is to attract greater funding for biodiversity research, and that "better marketing" by the systematics community would be a more constructive step towards this goal.



Window on research: children visiting Chicago's Field Museum watch Steve McCarroll prepare for display a pelvis bone from Sue, the most complete *Tyrannosaurus rex* ever discovered.

Raising the relevance to outside needs

Museums may be increasingly asserting the unique value of their traditional collections-based research. But their directors appear keenly aware that museum research can no longer remain an end in itself, and must be increasingly geared to answering questions of contemporary relevance — in particular about the management of the environment and biodiversity.

Two manifestations of this are a new emphasis on multidisciplinary research, and greater willingness to collaborate with external researchers. “The future of all natural history museums will depend on the strategic alliances we are able to build with other organizations to increase our reach and broaden our impact,” says Peter Crane, director of the Field Museum in Chicago. “We will need to develop an institutional culture that looks outwards rather than inwards.”

The Field Museum is bringing together scientists from different departments around shared facilities, for example in molecular biology. Similar proposals are contained in reforms submitted to the French government by Henry de Lumley, director of the Muséum National d’Histoire Naturelle in Paris. The museum’s 26 laboratories would be regrouped in ‘institutes’ built around particular research themes and collections — such as ecology and the management of biodiversity — with each sharing equipment and facilities.

At the Natural History Museum in London, reforms to encourage staff to seek research collaboration, both within the museum and in the wider community, have been under way for a decade (see opposite). Competition for external grants, for example, has become an important factor in internal promotion.

The museum’s research was previously too “cut off” from the outside world, argues Neil Chalmers, the museum’s director. He says he felt when he took over in 1989 that there was a need to have the museum’s unique research product — taxonomy — better appreciated outside the museum.

Paul Henderson, the museum’s director of science, sees the forging of external alliances as crucial to the institution’s future. He would like to see a graduate school developed between the museum and the University of London, for example, in much the same way as the American Museum of Natural History in New York has long had a productive link with Columbia University.

More broadly, managers at the London museum increasingly see its collection of 68 million specimens as a sort of collateral for attracting funds and staff for studies in such areas as biodiversity, conservation, agriculture, waste management, water treatment, and mineral exploitation.

Promoting the value of its collections for addressing contemporary problems, such as in biodiversity and the environment, has

become the focus of the museum’s development strategy. Chalmers defends himself against accusations that this amounts to chasing fashion at the expense of long-term interests by arguing that it reinforces the relevance of the collections in the eyes of funders.

A similar pragmatism now permeates many regional natural history museums, particularly in the United States. The Field Museum, for example, has become heavily involved in local issues, bringing together more than 40 state and local agencies and associations to form a conservation network known as the Chicago Wilderness Initiative.

“The funding opportunities are local,” says Edward Theriot, director of the Texas Memorial Museum in Austin, arguing that the central question facing all museums is one of “relevance”. Although scientific meetings may be preoccupied with global biodiversity, the museum is having to focus its biodiversity efforts on problems relevant to Texas, he says.

Similarly, Michael Hager, director of the San Diego Museum of Natural History, says the museum has completely refocused its activities over the past decade from global issues to biodiversity and environmental issues in California and Mexico. “The world kind of opened up for us, with funding becoming readily available,” says Hager, adding that ten years ago the museum was close to closing its doors.

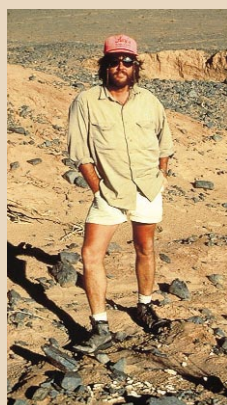
Museum world gears up to the ‘big bang’ of informatics

“What would Linnaeus have invented if he had lived in the twenty-first century and grown up with the Internet?” The question is posed in the prologue to a plan to create and link biodiversity databases around the world to provide a computer-based infrastructure similar to that which exists for molecular biology and genome research.

The proposal for a Global Biodiversity Information Facility (GBIF) has been put forward by the Working Group on Biological Informatics of the Megascience Forum of the Organization for Economic Cooperation and Development (OECD).

It is by far the largest of several initiatives to make collections-based research and information available to a wider audience by digitizing it and putting it on networked databases. It can then be integrated with information and maps on the ecology and biogeography of species, to provide sophisticated tools for the management of biodiversity and ecosystems.

The basic goal of the proposal is to create a web homepage for every species on the planet, beginning with existing data, and



Novacek: faster progress needed on collaboration.

later integrating data from projects aimed at creating a complete inventory of the world’s species, such as Systematics Agenda 2000. Users will eventually be able to verify the scientific name, status and classification of any species via a click on a web page.

The proposed multi-million dollar, 15-year programme is likely to be submitted for initial funding to OECD ministers next year. If approved, it would probably be brought under the umbrella of the United Nations Biodiversity Convention. “An immense amount of support is growing for GBIF,” says Steve Blackmore, keeper of botany at London’s Natural History Museum.

Many museum directors agree that digitization of the world’s collections, and their inclusion on web-based databases, is the way forward. Neil Chalmers, director of London’s Natural History Museum, argues that exploiting the opportunities offered by networked databases is essential to “translating research-based information in a form that is accessible and useful”.

“The major challenge now is to mobilize information on collections with geographical information system databases, to map flora and fauna in underexplored areas or for underexplored groups,” says Pieter Baas, chairman of Diversitas’s working group on creating a species inventory of the planet (see page 116).

An international initiative is imperative because the costs of digitalization are beyond the budgets of individual institutions. “Digitalizing the collections is essential,” says Ghilleen Prance, director of the Royal Botanic Gardens, Kew, in London. “But none of the big museums can afford to do it in a serious way — we are only doing little bits.”

At first, GBIF is expected to be selective to

Chalmers' choice: make cuts or go under

Few museums have witnessed as profound a transformation in recent years as the Natural History Museum in London has under the directorship of Neil Chalmers.

When he took up the post in 1988, the museum was struggling to make ends meet and almost totally dependent on dwindling public funds, all but two per cent of which were eaten up by salaries. But in the last financial year the museum generated £18.6 million (US\$31 million) of its £46.3 million turnover itself — just over 40 per cent.

Chalmers' actions have been controversial. They involved shedding or redeploying about 50 of the 290 scientific, curatorial and technical support posts, and refocusing research on fewer areas, putting some collections on a 'care-and-maintenance' footing (see *Nature* 344, 805; 1990).

Many experts in fields considered to be outside the museum's core interests have taken early retirement. Ten years on, there are now 270 staff funded by the museum, with a further 55 postdoctoral fellows and other staff paid by external agencies.

The link between museum research and the health of taxonomy and systematics as a discipline is a strong one. When the museum announced the job cuts in 1990, the then president of the Royal Society, Sir George Porter, received more than 800 letters, "all expressing concern that the great collections of the Natural History Museum and the research associated with them were in danger of neglect or



Chalmers: the results of museum research must be made public more speedily and effectively.

extinction through lack of support".

As a result, the House of Lords Committee on Science and Technology set up an inquiry into the state of systematic biology in the United Kingdom. Their recommendations, published in January 1992, included a modest increase in funding — £5 million

over five years — for systematics and taxonomy (see *Nature* 355, 488; 1992).

By a fortuitous coincidence, the debate on the committee's report in the House of Lords came soon after the United Nations Earth Summit in Rio de Janeiro. At the summit meeting, the then Prime Minister, John Major, announced the Darwin Initiative, to provide modest sums — £6 million to date — for systematics. Under the initiative, staff from museums and botanic gardens train taxonomists from developing countries.

Much of the resistance of researchers to the changes at the museum stemmed from opposition to the disbanding of teams of collections-based researchers on certain groups of species. The cuts have undoubtedly resulted in the irretrievable loss of a wealth of expertise which will have repercussions on the museum's activities for decades.

The cuts have been compensated for to some extent by the recruitment of around 100 university-style postdoctoral researchers financed by external agencies, such as the Natural Environment Research Council.

Two positive outcomes of the influx of outside scientists have been heightened external awareness of the museum's activities, and a greater 'footprint' in peer-reviewed journals, according to the director of science, Paul Henderson.

A further boost to researcher mobility has come from the recent award of a ECU680,000 (US\$750,000) grant from the European Union's Framework research programme to fund researchers from other European countries to visit the museum for up to three months at a time.

But financial austerity has forced the museum to abandon its scheme for funding undergraduates to spend the summer doing curatorial work, closing an avenue that in the past opened up careers for many scientists.

Chalmers defends his tough actions on the grounds that the urgency of the situation made drastic action necessary and that otherwise "we would have gone into a long and painful decline". Exhibitions would have been lacklustre, he argues, staff would have left and not been replaced, while the science would have become fragmented and unsupported, with no focus or plan.

According to Chalmers, a crucial challenge to museums in meeting real world needs will be to make collections-based research and information available in a form and at a speed — "unfamiliar to many who work in museums" — that makes them accessible to a wide range of users.

Chalmers argues that this requires researchers to communicate data "in brief and rapid reports," rather than in scholarly monographs, while taking full advantage of networked databases. □

keep costs down, focusing only on the most endangered or ecologically important groups. Given the sheer size of the world's collections, "you are talking millions, even if you only spend a few dollars per beetle," points out Stephan Michalowski, executive secretary of the OECD Megascience Forum.

Co-ordinating existing digitalization programmes is also a goal of the Consortium of European Taxonomic Facilities, a grouping of the major European museums. Similarly, an informal consortium of a dozen leading US museums is discussing co-ordinating their systematics work, including digitalization.

"We've made some progress, but it hasn't been enough," says Michael Novacek, provost of the American Museum of Natural History in New York. "Museums have had difficulty joining together to work on things because a lot of them have had to contend with challenges at home," he adds.

Discussions are also under way to integrate a UK-led programme, Species 2000, as the core of the early phases of GBIF. Species 2000 was launched in 1994 by the

International Union of Biological Sciences, in co-operation with the Committee on Data for Science and Technology and the International Union of Microbiological Societies.

The initiative has set its sights lower than GBIF, and would focus on bringing together databases of only existing described species to form a shared resource, while stopping short of digitalizing the specimens. Work is already under way to enter 40 per cent of the world's described species. Much of the information is coming from major museums, particularly in Europe.

Frank Bisby, professor of botany at the University of Reading in the United Kingdom and the chairman of the project, says that "pragmatism and quality" are its key criteria.

"Why haven't the taxonomists done this before?" asks Bisby. "At the moment there is gross inefficiency, with information distributed throughout countries and institutions." What we are planning is "desktop delivery of the world's existing knowledge on biodiversity".

Rescue plan needed for taxonomy

Sir — Following the House of Lords select committee report on systematic biology research¹, an important report on taxonomy was published by the UK Natural Environment Research Council (NERC)². This coincided with an article in *Nature* on the relentless decline in the number of posts in taxonomy in Britain³. This decline continues and applies to most of Europe and beyond^{4,5}. [As discussed elsewhere in this issue, see Briefing, pages 115–119.]

The NERC report identified the largest problem to be alpha (inventory) taxonomy (concerned with the recognition and description of species), as it estimated that only about 10 per cent of the species on Earth are known to science. The only parts of alpha taxonomy that have significantly benefited from the British government's most welcome Darwin Initiative are those programmes encouraging the production of user-friendly identification keys. Under the Darwin Initiative a British entomologist may be funded to visit an overseas country for an advisory or training mission or an overseas person may be funded to visit Britain for training⁶. However, a British entomologist will not be funded to carry out alpha taxonomy research on exotic species.

The British government's funding of the

UK Systematics Forum has done much to publicize the plight of taxonomy. So far not a single new post in alpha taxonomy has resulted from the forum's otherwise admirable efforts. Indeed it is now clear that alpha taxonomy will only return to UK institutions of higher education if ring-fenced money is allocated for the purpose⁷. The only practical suggestion that has so far been proposed to achieve this is the establishment of an Endowment Fund for Alpha Taxonomy¹. This suggestion was followed by the proposal that National Lottery money be used for the purpose. This proposal was evaded by the last government⁸ and the present government only responded when its failure to reply was publicized^{8,9}.

This publicity has now prompted a letter from Whitehall saying that the government is considering a National Endowment for Science, Technology and the Arts (NESTA), funded from the lottery. This could provide a source of funds for the proposed endowment fund for taxonomy. Whether NESTA will rise to the challenge will partly depend on the scientific community being able to persuade it of the urgent necessity to revive alpha taxonomy. Up to now the scientific community has tended to eclipse its appeal for more funding for taxonomy by

asking for increased funding for more glamorous fields of science.

It remains to be seen whether the present government, like the last, is unconcerned at the continuing decline of alpha taxonomy, despite political rhetoric in support of the need for biodiversity research. The fact that such research is inescapably grounded in alpha taxonomy eludes many politicians.

Will Britain's scientific community care if alpha taxonomy continues to become primarily the preserve of enthusiastic amateurs, most of whom are only interested in the species of their own country? The taxa currently requiring the greatest amount of research are precisely those for which a global perspective is essential.

Henry Disney

University Museum of Zoology,

Downing Street, Cambridge CB2 3EJ, UK

1. House of Lords Paper 22-II, 305 (London, 1991).

2. *Evolution and Biodiversity: The New Taxonomy* (NERC, Swindon, 1992).

3. Gaston, K. J. & May, R. M. *Nature* **356**, 281–282 (1992).

4. Blackmore, S. & Cutler, D. (eds) *Systematics Agenda 2000: The Challenge for Europe* (Linnean Society, London, 1996).

5. Black, C. C. (ed.) *Loss of Biological Diversity: A Global Crisis* (National Science Board, Washington, 1989).

6. Disney, R. H. L. *Sci. & Publ. Affairs* Autumn, 4–5 (1996).

7. Joysey, K. A. *FSC Mag.* **10**, 2 (1995).

8. Disney, R. H. L. *Biologist* **45**, 6 (1998).

9. Moore, P. G. *Biologist* **44**, 480 (1997).

Untenured in Toronto

Sir — I would like to respond to your News article "Fight heats up over Toronto racial discrimination claim", which gives an incomplete account of the facts of the case and misrepresents some of the issues in this dispute (*Nature* **392**, 638; 1998).

Kin-Yip Chun was one of many unsuccessful applicants for the two tenure-track positions filled in geophysics over the past decade by the Department of Physics at the University of Toronto. Your report gives credence to a claim that the two candidates actually hired "were less qualified than he was", a statement that would, among others, surprise the British university that had previously placed our first appointee as chair of its geophysics department.

Chun was hired by the Department of Physics in 1985 as a grant-funded research associate for a limited term of two years. Your statement that Chun "was told there were no funds to pay him" is untrue; this research associateship was fully funded.

Chun was employed as a research associate for his nine years with the university. In 1989, he was awarded a status-only appointment as an assistant professor so he could act as a principal investigator and a supervisor for graduate students. Although

he claims that "he was obliged to work as a professor", his academic duties were modest.

In December 1992, Chun insisted on continuing his research, but refused to sign a renewal of his contract. In November 1994, the university was forced to seek a campus police escort for Chun after he refused to consider all employment options offered to him and refused to vacate his office. Your article states that "no reason was given at the time". This is again false.

During his period as a research associate, Chun competed in 1987 and 1992 for two tenure-stream positions in the Department of Physics. Although you quote his claim that he "had been passed over for tenure", in fact he never held a position eligible for tenure. After the 1992 competition, Chun alleged that racial bias was a factor in his not being offered a tenure-stream position. To respond to these allegations, the university conducted several investigations. None of these, including one performed by vice-dean Cecil Yip, found any evidence to substantiate racial bias.

Finally, although you note the statement in Yip's report that the 1987 appointment could hypothetically have been made to "keep Chun out of the job", you omit to add that Yip found no evidence to support this.

The accusation of exploitation develops

from the blurring of roles that Chun faced once he was given the opportunity to be a principal investigator in seismological research in 1989. Yip's finding that Chun was placed in an exploitative situation has been accepted by the university. Consequently, the university has made a series of offers to Chun intended to remedy any harm that its actions may have had on his research and career.

Chun so far has rejected all these offers and insists that the only remedy is an academic appointment for a period of time equivalent to a *de facto* tenure-stream position. His demand contravenes the university's regulations on tenure, as well as the Canadian Association of University Teachers and the university faculty association's strongly held views.

Chun has more recently relaxed this demand, but other issues still remain unresolved. The University of Toronto continues to seek a negotiated settlement.

Pekka K. Sinervo (Chair)

Department of Physics, University of Toronto,
Toronto, Ontario, Canada M5S 1A7

We accept the specific corrections of statements made by *Nature* in the article, and apologize for any misleading impression that those errors may have caused. — Editor, *Nature*

Familiarity breeds cooperation

Laurent Keller and H. Kern Reeve

Many theoretical models have been developed to study the conditions under which unrelated individuals should cooperate or not cooperate. But such behaviour is rarely 'all or nothing', and new mathematical models allow the optimal level of cooperation to be determined.

Acts of giving, receiving and repaying permeate various aspects of human life and, similarly, the lives of many animals. However, evolutionary theory predicts that reciprocally altruistic cooperation should be rare in the animal kingdom — natural selection should tend to favour selfish, uncooperative behaviour. So what are the special conditions required for reciprocal altruism to occur? And what is the optimal level of altruism that each partner should show? New mathematical models reported by Roberts and Sherratt¹ on page 175 of this issue show that cooperation can spread more readily than previously thought. Moreover, cooperation should increase over the history of a partnership when organisms repeatedly interact, and when they can flexibly respond in a continuous manner to the level of cooperation that was previously shown by their partners.

Theoretical work on the evolution of altruism between unrelated individuals has been dominated by the 'prisoner's dilemma' model. In an encounter, each of two individuals can choose either to cooperate or not to cooperate ('defect' in trade jargon). If they both cooperate, both do better than if they had both defected. But if one player defects while the other cooperates, the defector gets the highest reward and the cooperator gets nothing, or even suffers a net cost. Clearly, if players meet only once the best option is to defect. However, if the players meet often enough, cooperation can be favoured². Computer tournaments between different strategies found that the most effective strategy was 'tit-for-tat', whereby players cooperate on the first move and thereafter repeat the opponent's previous move. In this way, they can punish defection but resume cooperation once the opponent does³.

Over the past 15 years there have been hundreds of theoretical papers aimed at working out which strategy would beat tit-for-tat under a range of conditions — for example, when individuals occasionally misidentify what opponents have done during the previous move⁴. Although of intrinsic theoretical interest, these hypothetical contests have been somewhat frustrating for empiricists, who have found it difficult to document the prisoner's dilemma array

of fitness payoffs (that is, reproductive consequences for cooperating and defecting) in real biological systems^{5,6}. In particular, a biologically unrealistic limitation of the conventional prisoner's dilemma is that it assumes that players can choose only between all-out cooperation and utter defection. But it is clear that cooperation is rarely an 'all or nothing' behaviour (Fig. 1).

Roberts and Sherratt¹ allowed for variable investment by each partner — with zero investment corresponding to defection — and analysed what would be the best strategy to adopt. Their computer simulations showed that, among several strategies compared, the most effective was 'raise the stake' (RTS). This strategy raises its investment with individuals who have matched or bettered their partner's last move. RTS was able to spread in the population until fixed when in competition with strategies which, for example, never invest, match a partner's investment, undercut a partner, or cheat probabilistically. Under most conditions, RTS was also successful in resisting sub-

sequent invasion by any of these defector strategies. The success of RTS stems from the fact that it is resistant to exploitation by defector strategies, while making the most of cooperative opportunities. In short, the success of RTS is due to the same features (being nice, retaliating and forgiving) that make tit-for-tat so successful in the discrete model of the prisoner's dilemma.

Perhaps the most important prediction to come from RTS is that individuals should invest relatively little during their first encounters with new partners, and then become increasingly cooperative with them. By 'testing the water', cooperators limit their losses but capitalize on cooperative opportunities with individuals who are prepared to reciprocate. This makes sense. Are you more likely to lend £100 to a friend or to a stranger? The idea of earlier encounters being used to build trust has also been proposed by May⁷. He suggested that partners may eventually undertake an enterprise with a large reward — say, robbing a bank — once they have built up enough trust during previous encounters. Unfortunately, empirical studies recording changes in the degree of cooperation during successive interactions are scarce. Experiments with humans showed that people become increasingly cooperative during four successive interactions⁸. But, in that study, the options were only discrete cooperation or defection, as in the original prisoner's dilemma game.

Roberts and Sherratt¹ also suggest that potlatch (a custom among certain North American Indians whereby property is given away or destroyed) may represent altruistic escalation. However, some caution is needed here as potlatches may serve as honest signals



Figure 1 You scratch my back and I'll scratch yours — grooming by Barbary macaques on Gibraltar. Roberts and Sherratt¹ have devised new models for cooperation that take into account the fact that, in biological situations, the choice is not always as clear-cut as whether or not to cooperate. One example of this is the amount of time that macaques who cooperate by grooming one another invest in this altruistic act.

R. D. MARTIN, AIM, ZÜRICH

of social roles, wealth and status⁹, potentially providing other benefits to the 'greatest potlatches'. The significance of gift-giving has been discussed by economists, many of whom believe that gifts may serve as signals of a person's intention about future investment in relationships.

Experiments with children showed that they give more to friends than non-friends in some situations (as one would expect if friendship is the outcome of trust building up between partners). In one study¹⁰, for example, children gave more trinkets to friends than to non-friends. But they tended to decrease the amount given to friends after receiving many trinkets from them, and to increase the number given after receiving few. The reverse occurred with neutral or disliked partners. When little was received from a friend it was interpreted as punishment for previous selfishness, and a lot was returned. When a lot was received from a non-friend, it was interpreted as an invitation to friendship (and interpersonal gain) and a lot was given in return¹⁰. These experiments show that although levels of cooperation genuinely correlate with friendship, in accordance with the Roberts and Sherratt models, other factors such as reconciliation after perceived punishment and honest signalling of willingness to cooperate in the future may also affect the magnitude of dispensed cooperation, at least in humans. We suspect that such cases will be explained by models of cooperation that also take into account the preferential assortment of co-operators according to cooperative history¹¹

or need (individuals having the same status or age, for instance).

An important and welcome feature of Roberts and Sherratt's new models is that they make predictions that can be easily tested. For example, several systems exist in which changes in the levels of investment between partners can be followed over time. It is to be hoped that Roberts and Sherratt's models (and future models, similarly inspired by the details of how animals actually behave) will forge a stronger link between theoretical and empirical studies — allowing theoreticians and empiricists to build trust and cooperate to a greater extent in testing evolutionary models of cooperation. □

Laurent Keller is in the Department of Zoology and Animal Ecology, Bâtiment de Biologie, University of Lausanne, 1015 Lausanne, Switzerland.

e-mail: laurent.keller@izea.unil.ch

H. Kern Reeve is in the Section of Neurobiology and Behavior, Seeley G. Mudd Hall, Cornell University, Ithaca, New York 14853, USA.

e-mail: hkr1@cornell.edu

1. Roberts, G. & Sherratt, T. N. *Nature* **394**, 175–179 (1998).
2. Axelrod, R. & Hamilton, W. D. *Science* **211**, 1390–1396 (1981).
3. Axelrod, R. *The Evolution of Cooperation* (Basic Books, New York, 1984).
4. Dugatkin, L. A. *Cooperation Among Animals. An Evolutionary Perspective* (Oxford Univ. Press, New York, 1997).
5. Connor, R. C. *Trends Ecol. Evol.* **10**, 84–86 (1995).
6. Noe, R. *Anim. Behav.* **39**, 78–90 (1990).
7. May, R. M. *Nature* **327**, 15–27 (1987).
8. Wedekind, C. & Milinski, M. *Proc. Natl Acad. Sci. USA* **93**, 2686–2689 (1996).
9. Camerer, C. *Am. J. Sociol.* **94**, S180–S214 (1988).
10. Krebs, D. L. *Psychol. Bull.* **73**, 258–302 (1970).
11. Wilson, D. S. & Dugatkin, L. A. *D. Am. Nat.* **149**, 336–351 (1997).

that have had recent mergers. But the universal profile is still a tempting target for observers to shoot at.

Tyson *et al.*¹ took their shots using the Hubble Space Telescope. Although the telescope was pointed at the cluster of galaxies 0024+16, at a moderate redshift of 0.39, the real target was the backdrop of faint galaxies lying behind the cluster. Light from these galaxies is deflected by the gravitational field of the intervening cluster, as predicted by Einstein's theory of General Relativity, and the deflection warps the images of the galaxies. The effect is weak in the outer parts of the cluster but becomes strong in the centre, where the density is highest and the gravitational potential well is deepest.

Strong lenses such as 0024+16 create multiple images of background objects, and the appearance of such a multiply-imaged system is sensitive to the central structure of the cluster's potential well. Potential and mass density are related by Poisson's equation, so the image structure can be 'inverted' by numerical methods to constrain how mass is distributed within the cluster core.

Recent simulations⁴ have revealed central potential wells that are 'cuspiest' (more sharply peaked in the centre) than the universal type described earlier. But to the theorists' dismay, Tyson *et al.* infer a core structure for 0024+16 that not only differs from the universal prediction, but also differs in the direction opposite to that seen by the new simulations. Their cluster reconstruction has a much flatter central peak (Fig. 1).

What to make of this unpleasantness? Do the simulators need to reprogramme? Is this a hint of new physics? Or is it merely accidental? Before getting too excited, we should keep in mind that the size of the discrepancy is small: the difference in core mass between theory and the new observation is only 1% of the total mass inferred for the cluster.

Simulators hammering at the structure of clusters have come a long way since the pioneering numerical work of Peebles⁵ in 1970. A recent comparison of 13 simulation codes shows agreement in the dark-matter density structure at the 5% level. However, a fair criticism of the numerical models is that the very highest-resolution versions, which probe the core region best, do not take into account the gravitational coupling of the dark matter to ordinary 'baryonic' matter — the galaxies and intra-cluster plasma. Overall, these contribute nearly 20% of the total mass budget. It is conceivable, at least on energetic grounds, that dynamical mechanisms could pump enough heat into the dark matter to alter the structure of its inner 1%. Furthermore, the predicted 'universal mass profile' strictly applies only to the dark-matter component, whereas the observations constrain the structure of all matter outside

Galaxy clusters

Well of darkness

August E. Evrard

Astronomers have gone to the well and come back perplexed. Galaxy clusters should contain large quantities of dark matter, trapped in their potential well, and we expect that matter to be sharply concentrated near the centre. But using a new method for interpreting the effects of gravitational lensing, Tyson *et al.*¹ have inferred a much smoother and less centrally concentrated distribution of dark matter in one cluster. If this distribution is confirmed, and seen in other clusters, it is likely to produce insights into the mechanisms governing galaxy-cluster formation. A less likely, but more exciting, possibility, is that it is a clue to the very nature of dark matter.

The standard mechanism of large-scale structure formation is gravitational instability². Small irregularities in density grow with time, collecting matter into an evolving, web-like network. Within this large-scale network, collapsed, nearly spherical, quasi-equilibrium structures called 'haloes' form,

their internal gravitational attraction balanced by pressure from the random motions of their elements. As the Universe ages, collections of bound haloes merge, and an evolving hierarchy of collapsed structures is established. Clusters of galaxies mark the apex of this process — they are the biggest collapsed structures around.

Over the past decade, this has been scrutinized using numerical simulations of gravitational clustering. To produce anything like the current distribution of clusters, there must be a great deal of dark matter present as well as the luminous matter that we see. Most simulations predict similar internal structures for clusters; recent work³ even points to a 'universal profile' for the radial dark-matter density distribution — a simple function with one free parameter that is sensitive mainly to the underlying cosmology (the Universe's density and expansion rate). The main caveat is that these numerical studies neglect the 10–30% of clusters

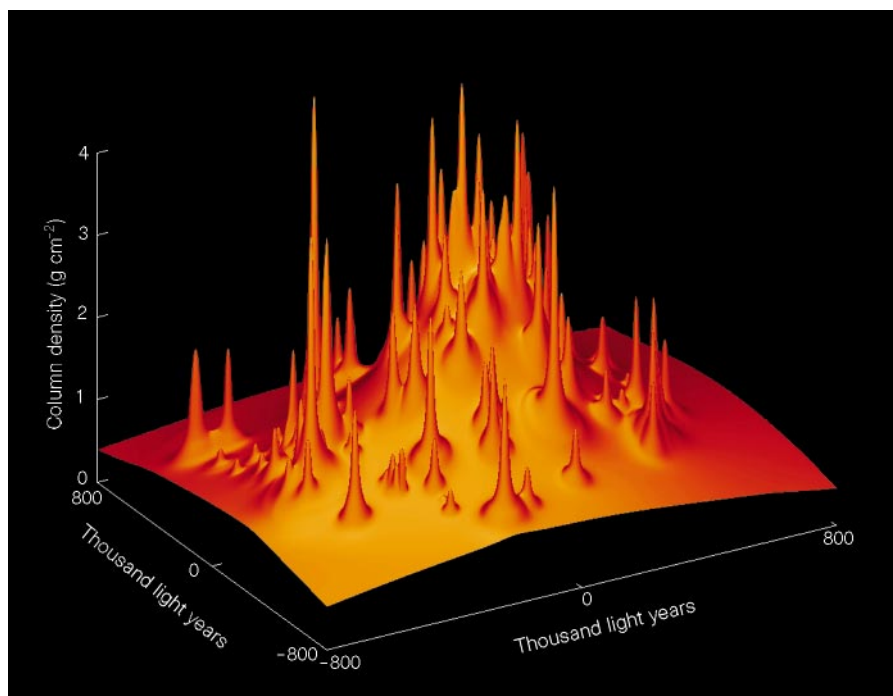


Figure 1 Profile of a lens. This is a projected density plot of the galaxy cluster CL0024+1654, inferred from the cluster's gravitational effect on the images of more distant objects. (The sharp peaks are added to mark single galaxies in the cluster.) The mass in the cluster is less centrally concentrated than theories predict. Is this just an unusual cluster, or could it be a sign of the existence of 'hot' dark matter?

galaxies, including the intracluster plasma. Because the plasma is known to be less centrally concentrated than the dark matter, its effect on the mass profile would tend to reconcile theory and experiment.

New fundamental physics is always more fun than new astrophysics as a solution to cosmic mysteries. A possible solution from this category is that the dark matter is not all cold. A hot, relativistic component would have too low a phase space density to create a very cuspy potential⁶, so it would lower the inner parts of the cluster density profile, perhaps enough to satisfy the observations⁷. The new evidence for flavour oscillations of atmospheric neutrinos from Super-Kamiokande⁸, discussed in these pages by Frank Wilczek⁹, may mean that neutrinos have enough mass to do this.

More mundane solutions exist, of course. Tyson and colleagues use a complex fitting procedure, summing the effects of many individual mass concentrations ('mascons'), and end up with a model that has 512 free parameters — a frighteningly large number. One might be suspicious of finding a global best fit in such a large parameter space. But the much larger number of independent pixels (3,800) in the multiple galaxy images provides the grist for this modelling mill; and the end result, that just two superimposed mascons contain 98% of the non-galaxy mass, has a ring of trustworthiness about it.

Finally, 0024+16 may just be a bad egg. Because of their immense gravitational fields, clusters are constantly attracting and

swallowing nearby cosmic debris. When smaller objects are absorbed they usually plunge near to the heart of the cluster while gravitational tides deform and ultimately dissolve them. If 0024+16 is in the process of digesting such a snack, that would invalidate comparison with simulation predictions that select against such objects.

What is ironic about this finding is a sense of inverted déjà vu. Only a few years ago, Tyson and co-workers stirred up the waters by revealing cluster cores that were *more* centrally cusped than the ageing, but still popular, theoretical model then in use. That finding helped spur the theoretical investigations that resulted in the universal density profile's enshrinement. Let's see: first time to the well came up too sharply cusped; second time, not cusped enough. One more time, perhaps, and it will come out just right. □

August E. Evrard is in the Physics Department, University of Michigan, 500 East University, Ann Arbor, Michigan 48109-1120, USA.
e-mail: evrard@umich.edu

1. Tyson, J. A., Kochanski, G. P. & Dell'Antonio, I. P. *Astrophys. J. Lett.* **498**, L107–L110 (1998).
2. Peebles, J. *Principles of Physical Cosmology* (Princeton Univ. Press, NJ, 1993).
3. Navarro, J. F., Frenk, C. S. & White, S. D. M. *Astrophys. J.* **490**, 493–508 (1996).
4. Moore, B., Governato, F., Quinn, T., Stadel, J. & Lake, G. *Astrophys. J. Lett.* **499**, L5–L10 (1998).
5. Peebles, J. *Astron. J.* **75**, 13–19 (1970).
6. Tremaine, S. & Gunn, J. E. *Phys. Rev. Lett.* **42**, 407–419 (1979).
7. Kofman, L., Klypin, A., Pogosyan, D. & Henry, J. P. *Astrophys. J.* **470**, 102–114 (1996).
8. <http://www-sk.icrr.u-tokyo.ac.jp/doc/sk/index.html>
9. Wilczek, F. *Nature* **391**, 123–124; **394**, 13–14 (1998).



100 YEARS AGO

In India, as elsewhere, the general doctrine of disease prevails that all abnormal and morbid states of body and mind are caused by *demons*, who are conceived either as attacking the body from without, or as temporarily entering the body of man. The consequence is that primitive medicine consists chiefly in chasing away or exorcising these hostile spirits. This is done, in the first instance, by *charms*. The spirit of the disease is addressed with coaxing words and implored to leave the body of the patient, or fierce imprecations are pronounced against him, to frighten him away. But these charms, powerful as they are (in fact, there is nothing more powerful to the primitive mind than the human word, the solemn blessing or curse), are yet not the only resource of the ancient physicians or magicians. From the earliest times people had become aware of the curative power of certain substances in nature, especially of herbs. ... The principle that *similia similibus curantur* prevails throughout the whole range of folk-medicine. Thus dropsy is cured by water.

From *Nature* 7 July 1898.

50 YEARS AGO

Previous methods do not appear to give any indication of how to distinguish between young and old Colorado beetles, except during the first few days following the emergence of the young beetle. I have found evidence that it is possible to separate young and old beetles for a period of at least fourteen days after the emergence of the young adult by means of colour changes in the membranous wings. This fact is of considerable practical importance and assistance to entomologists, especially in countries where the beetle has not yet become an endemic pest. ... The membranous wings of the fresh emergent are transparent and devoid of colour, while those of the mature beetle are red except for the apex and anal border of the wing. The wings remain colourless for at least four to five days after emergence, and afterwards develop a pink colour which gradually deepens and diffuses over the greater part of the wing surface ... until the pink colour has given place to a distinct red hue.

From *Nature* 10 July 1948.

Cell motility

Actin branches out

Laura M. Machesky and Michael Way

The ability of cells to move is essential for the development and survival of multicellular organisms, yet it can also be detrimental — when cancer cells migrate to distant parts of the body and colonize new tumours, for example. Cell movement is thought to be driven largely by the polymerization of actin monomers into filaments near the plasma membrane, and by the contractile forces of myosin motors found deeper in the cell. Although the mechanism of muscle contraction is fairly well understood, we still don't know much about the roles of actin and myosin in cell migration. But reports by Mullins *et al.*¹ in *Proceedings of the National Academy of Sciences* and Welch *et al.*² in *Science* reveal that a seven-protein complex is central in organizing actin structures. This complex contains the actin-related proteins Arp2 and Arp3, and has been thrust into the spotlight because it may initiate and organize protrusion of an actin-based dynamic meshwork into thin sheets (lamellipodia) or spikes (filopodia).

First identified in *Acanthamoeba*³, the Arp2/3 complex is conserved from yeast to mammals. In fibroblasts and amoebas, this complex is localized at the leading edge of cells — the ideal position to initiate the polymerization of actin that is required for cell motility. However, the Arp2/3 complex is not found in the more stable actin-myosin bundles that are normally associated with adhesion and contraction^{4–6}. In budding yeast, the Arp2/3 complex is required for motility of the cortical actin patches (dynamic, actin-rich structures that may be analogous to mammalian cortical actin) in which it is found^{7,8}. It is also the only known host-encoded factor that can induce actin polymerization on the surface of the motile intracellular pathogen *Listeria monocytogenes*⁹.

Mullins *et al.*¹ and Welch *et al.*² now explain why the Arp2/3 complex is associated with dynamic actin structures. They find that it seeds, or 'nucleates', actin filaments to polymerize from their barbed (fast-growing) ends (Fig. 1). This has led Mullins *et al.* to propose an attractive model for the mechanism of actin polymerization at the leading edge of the cell (Fig. 2). In this model, an actin dimer elongates at its barbed end while being capped at its pointed (slow-growing) end by the Arp2/3 complex (Fig. 2). The authors call this mechanism 'dendritic nucleation' because the *in vitro* data¹ indicate that new filaments could be nucleated off the sides of old ones, as well as being made from scratch (inset in Fig. 3, overleaf). Interestingly, Svitkina *et al.*¹⁰ have observed similar actin branches in lamellipodia of fibroblasts (Fig.

3) and motile keratocytes. These authors proposed that branches might be nucleated by the Arp2/3 complex and drive protrusion of the lamellipodia.

Pointed-end capping and branching by the Arp2/3 complex add a new dimension to the way we think about actin dynamics at the leading edge of motile cells. Previously, the barbed ends of actin filaments near the plasma membrane were thought to be blocked by the binding of barbed-end capping proteins

until the caps dissociated in response to extracellular signals¹¹. Once free, these barbed ends then grew rapidly until they were re-capped. In this model, the pointed ends of the actin filament were always assumed to be free. But the new model of dendritic nucleation suggests that actin polymerization occurs not only by uncapping, but also by Arp2/3-driven nucleation of filaments with free barbed ends and capped pointed ends (Fig. 2).

The interplay between nucleation and uncapping, as well as the branching and bundling activities of the Arp2/3 complex, may help to explain the relationship between subtly different structures at the leading edge of cells. Motile cells such as amoebas,

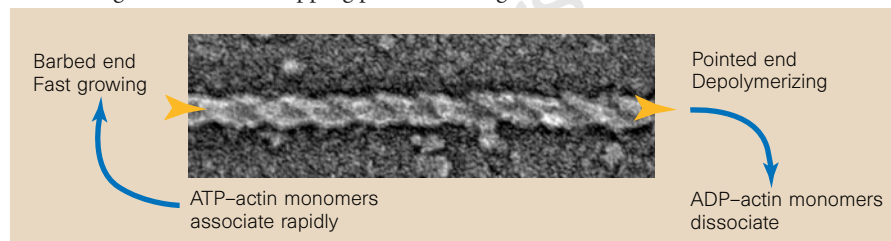


Figure 1 Structure of an actin filament. The filament has polarity, conferred by its structure and by the bound adenine nucleotide. The filament shown is decorated with the S1 sub-fragment of myosin II, which binds in a polarized fashion (yellow arrowheads). The barbed side of the arrow represents the fast-growing end of the filament, whereas the point of the arrowhead represents the depolymerizing end. Mullins *et al.*¹ have found that the Arp2/3 complex caps the pointed end with high affinity. Bound ATP is hydrolysed randomly, so the older regions of the filament (nearer the pointed end) have mostly ADP bound. (Reprinted in modified form with permission from Svitkina and Borisov¹⁵.)

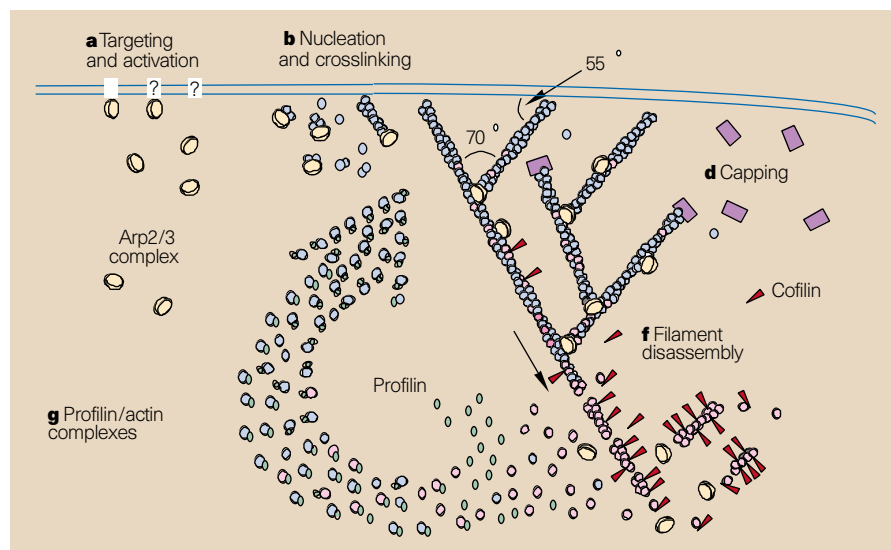


Figure 2 Dendritic-nucleation model for protrusion of lamellipodia. a, The Arp2/3 complex is targeted to sites where the cell has received a signal to form a lamellipodium. b, The newly activated Arp2/3 complex nucleates actin filaments off the sides of pre-existing filaments to form a branched network of actin, which drives protrusion of the leading edge. c, Actin filaments, capped at their pointed ends by the Arp2/3 complex, can elongate transiently and rapidly at their free barbed ends. d, The rapid growth is terminated by capping of the barbed ends, imposing an additional regulation on the elongating filaments. e, As ATP in the actin filaments is hydrolysed to ADP, the filaments become more susceptible to being severed by cofilin. f, Severing promotes disassembly, mainly in 'older' regions of the filaments, closer to the cell body. g, The actin-monomer-binding protein profilin could then bind to the newly released ADP-actin monomers, promoting nucleotide exchange and recharging them for polymerization into new filaments. (Reprinted in modified form from Mullins *et al.*¹.)

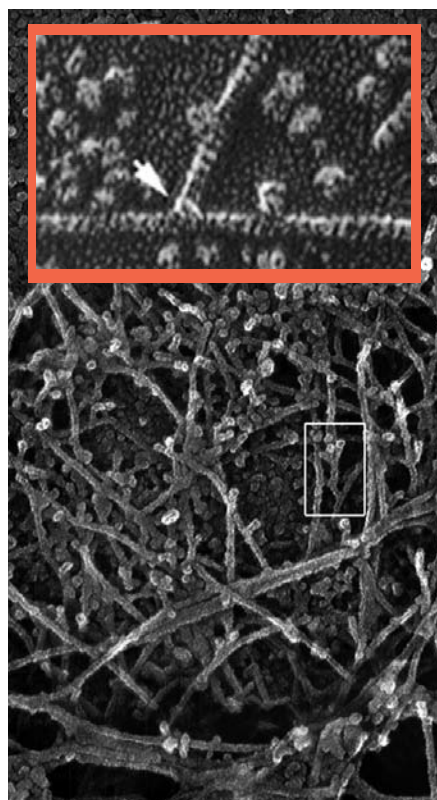


Figure 3 Electron micrograph showing actin filaments and bundles in the lamellipodium of a fibroblast. The inset shows Arp2/3 complex mixed with gelsolin-capped actin filaments *in vitro*. The branch points both *in vitro* (inset) and *in vivo* (white box) have a similar fixed angle and seem to contain one Arp2/3 complex at the junction.

leukocytes and growing neurons assemble both filopodia and lamellipodia at their leading edges. Filopodia are thought to sense the environment and direct polarized movement, whereas the actin-rich lamellipodia protrude from the cell in the direction of movement. If the Arp2/3 complex regulates the interplay between orthogonal (lamellipodia) and parallel (filopodia) networks, as well as controlling rates of actin assembly, it could be a control centre for actin dynamics and organization at the leading edge. This may be similar to the way in which a focal adhesion is thought to control both the sensing and mechanics of cell adhesion.

The dendritic-nucleation model also means that we must revise our ideas about turnover of actin filaments in motile cells. The density of actin filaments is highest at the leading edge of a lamellipodium, gradually decreasing towards the middle of the cell^{10,12}. This was previously attributed to disassembly of the free pointed ends nearer the centre at a fairly constant rate, as well as the action of cofilin — a protein that depolymerizes and severs actin filaments¹³. But if the Arp2/3 complex caps all of the pointed ends, these ends can only be disassembled if the actin filament is first severed by proteins

such as cofilin. Because of ATP hydrolysis, older regions of an actin filament tend to contain bound ADP, so they are more susceptible to depolymerization or being severed by cofilin. ATP hydrolysis on actin could also mediate dissociation of the Arp2/3 complex from pointed ends, triggering rapid depolymerization. This dynamic disassembly could be important for cellular retractions or rapid changes in direction.

Few aspects of actin dynamics seem to be untouched by the activities of the Arp2/3 complex. But several questions remain. How do actin filaments turn over in the cell, given the high concentrations of both barbed- and pointed-end capping proteins? How does cofilin cooperate with the Arp2/3 complex to promote disassembly? And how is the Arp2/3 complex regulated, both temporally and spatially? Welch *et al.*² may have already set us crawling in the right direction by their observation that the amino-terminal half of the ActA protein from *Listeria* causes a 50-fold increase in the nucleation activity of the Arp2/3 complex. ActA is the only bacterially encoded protein that is needed for actin polymerization on the surface of *Listeria*¹⁴, and the authors suggest that some cellular proteins may be functionally homologous to the amino terminus of ActA. These proteins could regulate the activity of the Arp2/3 complex in an analogous fashion. So, the race is now

on to find what regulates the Arp2/3 complex in the cell and to understand how this complex impinges on cell motility through its various activities. □

Laura M. Machesky is at the MRC Laboratory for Molecular Cell Biology, Gower Street, London WC1E 6BT, UK.

e-mail: dmcblam@ucl.ac.uk

Michael Way is at the European Molecular Biology Laboratory, Meyerhofstrasse 1, D-69117 Heidelberg, Germany.

e-mail: way@embl-heidelberg.de

1. Mullins, R. D., Heuser, J. A. & Pollard, T. D. *Proc. Natl Acad. Sci. USA* **95**, 6181–6186 (1998).
2. Welch, M. D., Rosenblatt, J., Skoble, J., Portnoy, D. A. & Mitchison, T. J. *Science* **281**, 105–108 (1998).
3. Machesky, L. M., Atkinson, S. J., Ampe, C., Vandekerckhove, J. & Pollard, T. D. *J. Cell Biol.* **127**, 107–115 (1994).
4. Welch, M. D., DePace, A. H., Verma, S., Iwamatsu, A. & Mitchison, T. J. *J. Cell Biol.* **138**, 375–384 (1997).
5. Mullins, R. D., Stafford, W. F. & Pollard, T. D. *J. Cell Biol.* **136**, 331–343 (1997).
6. Machesky, L. M. *et al. Biochem. J.* **328**, 105–112 (1997).
7. Moreau, V., Madania, R., Martin, P. & Winsor, B. J. *Cell Biol.* **134**, 117–132 (1996).
8. Winter, D., Podtelejnikov, A. V., Mann, M. & Li, R. *Curr. Biol.* **7**, 519–529 (1997).
9. Welch, M. D., Iwamatsu, A. & Mitchison, T. J. *Nature* **385**, 265–269 (1997).
10. Svitkina, T. M., Verkhorvsky, A. B., McQuade, K. M. & Borisy, G. G. *J. Cell Biol.* **139**, 397–415 (1997).
11. Hartwig, J. H. *et al. Cell* **82**, 643–653 (1995).
12. Anderson, K. L., Wang, Y.-L. & Small, J. V. *J. Cell Biol.* **134**, 1209–1218 (1996).
13. Maciver, S. K. *Curr. Opin. Cell Biol.* **10**, 140–144 (1998).
14. Smith, G. A., Portnoy, D. A. & Theriot, J. A. *Mol. Microbiol.* **17**, 945–951 (1995).
15. Svitkina, T. M. & Borisy, G. G. *Methods Enzymol.* (in the press).

Superconductivity

Disorder in the ranks

R. J. Cava

By changing the dopant in a high-temperature superconductor, the transition temperature can be changed markedly. Why this happens has been obscure, but on page 157 of this issue¹, Attfield *et al.* show that it is due to lattice strains, caused by disorder in the cation size. This solves a ten-year-old puzzle — but leaves the question of the microscopic mechanism open.

In the early days of research into high-temperature superconductivity, many believed that a microscopic explanation based on the theory of low-temperature superconductivity was possible. In this picture, electron–phonon coupling binds electrons into ‘Cooper pairs’ below T_c , the superconducting transition temperature. (Phonons are quanta of lattice vibration.) The total pair momentum is unaffected by lattice interactions, and the pairs therefore flow without resistance.

The first nails in the electron–phonon coffin were the observations of superconductivity at 90 K in Y–Ba–Cu–O, and soon afterwards, of a small or non-existent effect of phonon frequency (changed by isotope

substitution) on T_c . So the argument has been made that high-temperature superconductivity is made possible specifically by the lack of strong coupling between the electronic state of the compounds and their crystal lattices (far from the normal case for first-row transition-metal oxides). Nevertheless, there is evidence that superconductivity and the underlying lattice are in fact linked — evidence ranging from observations of changes in some phonon modes at T_c to the observation of a sharp dip in T_c at a doping level of an eighth of a hole per copper in (La,Ba)₂CuO₄ and an accompanying change in the symmetry of the crystal structure. But is such coupling a manifestation of an essential interaction, or only an incidental one^{2,3}?

La₂CuO₄ has been a proving ground for ideas about high- T_c superconductivity. It consists of single CuO₂ layers sandwiched between double LaO layers, where dopant atoms can be partially substituted for lanthanum, contributing charge-carrying holes and so allowing control of the electronic properties. The atoms on the LaO layers only indirectly affect the CuO₂ layers, through the bonding of the large lanthanide and dopant



Figure 1 Super conduct: the line of soldiers has less size disorder than the marching band. In superconductors, the disorder is fixed by the sizes and mix of different cations.

ions to the oxygen in the CuO_2 sheet. Hole concentration and lattice size affect T_c . The choice of dopant used to attain the optimal hole concentration also affects T_c , but no convincing explanation for the variations has been proposed.

Now, Attfield *et al.*¹ show that, at a fixed hole concentration and lattice size, the degree of lattice-size disorder in the LaO layer has a powerful effect on T_c . Those lattices with minimal size disorder (that is, whose dopant ions have similar radii to the lanthanum ions) have the highest T_c ; those with bigger or smaller dopant ions, and therefore more disorder and stress in the lattice, have lower T_c . And this is in materials where the lattice is supposed by many to be irrelevant, except in providing the CuO_2 plane as a sort of flower-bed in which exotic electronic states can come to life.

Some readers may be left dissatisfied because the authors do not try to explain the observed effect with a microscopic model, but it seems wise to me. We are in the midst of what seem to be weekly developments and discussions about the superconducting mechanism in cuprates, and these data point to none of the options in a straightforward manner. Although the authors have observed a similarly strong effect of disorder on the magnetoresistance in manganites, it would be surprising if the underlying mechanism is the same in the two systems. The effect could be something trivial — such as the smearing out of the electronic density of states by disorder — or more subtle, perhaps reflecting the competition between how strongly the dopants localize their charge on nearby transition-metal-oxygen units, and how strongly such charges tend to delocalize.

This paper helps to rid me of an old demon. In early December 1986, a physicist colleague grabbed me in the hallway on his way to a seminar to be given by Koichi Kitazawa (who was visiting Bell Labs from the University of Tokyo) on a new kind of superconductor. He dragged me in, and I sat in the back row, feeling distinctly out of place —

superconductivity at the time being the domain of condensed-matter physicists and not solid-state chemists. Kitazawa had agreed to give an informal talk on Tokyo's confirmation of superconductivity at 28 K in the La-Ba-Cu-O chemical system, published only weeks before by the two Swiss scientists Georg Bednorz and Alex Muller. During the seminar, a telephone was brought into the room and Kitazawa called the lab back home (where it was 4 a.m.), to learn that they had just determined that the superconducting compound in the system was the now famous La_2CuO_4 structure, doped with barium.

I left the room after the seminar thinking that if barium was a good dopant, then certainly strontium should be better, being almost identical in size to lanthanum. With-

in a few days our group and others around the world had seen superconductivity at the then astounding temperature of 38 K in the strontium analogue $\text{La}_{1.8}\text{Sr}_{0.2}\text{CuO}_4$. The cuprate genie was out of the bottle. Although the idea seemed simple enough at the time, like many things about copper oxide superconductors to follow, we were soon to figure out that the story was by no means so simple. I have been wondering ever since why that small change in dopant should increase T_c so dramatically. Attfield and co-workers have now provided an answer.

One might find it astonishing that more than a decade would pass since the first observations of superconductivity in this family of compounds and the presentation of a unifying hypothesis explaining the observed variations in T_c . The delay is a testimony to the complexity of this still remarkable class of materials, especially the way they continue to shroud their true nature beneath layer upon layer of obscurity. It is also a testimony to the nature of the scientific process, in which some with just the right point of view see clearly what was invisible to many others. The delay makes the understanding all the more sweet. □

R. J. Cava is in the Department of Chemistry and the Princeton Materials Institute, Bowen Hall, 70 Prospect Avenue, Princeton, New Jersey 08540, USA.

e-mail: rcava@chemvax.princeton.edu

1. Attfield, J. P., Kharlanov, A. L. & McAllister, J. A. *Nature* **394**, 157–159 (1998).
2. Ginzburg, V. L. *Contemp. Phys.* **33**, 15–23 (1992).
3. Anderson, P. W. *The Theory of High Temperature Superconductivity* (Princeton Univ. Press, NJ, 1997).

Palaeontology

Ancient Australian arthropods

This little creature, some 3½ cm in length and probably a relative of present-day centipedes and millipedes (myriapods), was crawling around what is now Australia almost 400 million years ago. As G. D. Edgecombe reports elsewhere in this issue (*Nature* **394**, 172–175; 1998), this fossil and a younger one were identified in deposits that date to the Early and Mid Devonian, respectively, and plausibly become the oldest-known Australian land animals. That they were terrestrial is surmised from what seems to be a spiracle, the external opening of the tracheal system employed by air-breathing arthropods, of which myriapods are a constituent group.

In the Devonian, today's southern continents plus India were formed into the supercontinent Gondwanaland. Previous examples of terrestrial arthropods of Devonian age have come only from the other supercontinent of the time, Laurasia. Most notably, the new fossils are assigned to the genus *Maldybulakia* from Kazakhstan, and so greatly extend the



known geographical range of this type of myriapod. That begs the question of how the two places might have been connected in the past. Edgecombe points to a Lower Devonian reconstruction which links Kazakhstan and Australia by a land bridge through north China. Further specimens might be found there.

One hopes, however, that miracles of fossilization mean that other examples will be the remains of complete animals. These two have lost their heads.

Tim Lincoln

Neuroscience

Dancing dendrites

Frances A. Edwards

Dendritic spines are small, actin-rich structures, about 1 μm in diameter, that protrude from the sides of dendrites and receive most of the excitatory inputs onto neurons in the brain (Fig. 1). Reporting in *Neuron*, Fischer and colleagues¹ have used elegant video imaging to show that the position and shape of these dendritic spines change constantly, yet they vary little in cross-sectional area. Within two seconds, the position of the edge of a spine can move as far as 100 nm, while over two minutes it can move by more than 300 nm — up to 30% of the total width or length of the spine (Fig. 2).

Fischer *et al.* first transfected cultured hippocampal neurons with actin that was tagged with green fluorescent protein. Then, using videomicroscopy, they monitored any changes in the position of the actin on a second-to-second timescale. Each spine appeared like a tiny writhing sack, full of actin filaments struggling unsuccessfully to escape. The movement was clearly mediated by actin polymerization, as it was reversibly inhibited by cytochalasin, an agent that blocks this process. What does all this movement mean with regard to the function of dendritic spines in the brain? Is it an example of what happens functionally, or is it an artefact of cell culture? Could it be involved in synaptic plasticity, or development, or both? Whatever the answers, this demonstration of dynamic activity shows that substantial movement is undoubtedly something that the spines can do.

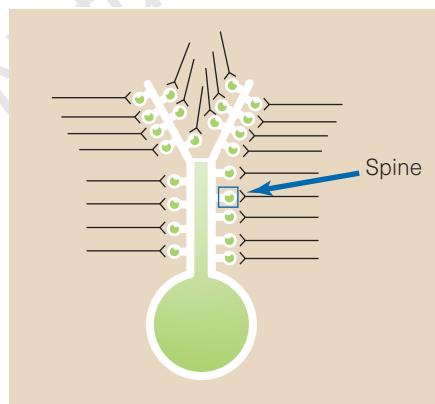


Figure 1 A typical spiny cell in the central nervous system. The small box outlines a dendritic spine, and the black forked lines represent incoming excitatory axons that release glutamate onto the spines. Note that a single pyramidal cell in the hippocampus typically features tens of thousands of spines all over its dendritic tree, although in some cell types few spines occur before the first branch of the dendritic tree.

There is a well-studied precedent for the involvement of actin-mediated structural changes in synaptic plasticity. Actin forms filamentous structures that, among other functions, mediate light adaptation by the formation of huge outgrowths (spinules) from the dendritic spine in synapses between horizontal cells and cones in fish retina (>100 nm in 30 min; refs 2, 3). It thus seems likely that the rapid movement in cultured hippocampal neurons, seen at high time-resolution in this study, stems from a physiological phenomenon, but it may not reflect what would occur in the brain — the physical interaction and anchoring of neurons to surrounding cells in the brain, in a three-dimensional array, is undoubtedly somewhat different from the two-dimensional array of a culture dish.

Neurons are tightly packed together in the brain, with glia (non-neuronal cells) filling every space and closely surrounding every synaptic contact. Presumably, this would restrict the possible movement of dendritic spines. In fact, if the millions of spines in the brain were all constantly moving to the extent seen in this study, the whole brain would become a jumbling mass of movement. So, although it seems very likely that actin grows and retracts in a way that would cause such movement in a loose array, this movement would presumably be much more restricted in the brain.

What is the significance of such dynamic changes in shape? It may be a developmental phenomenon. Thus, despite having successfully formed synaptic connections, the spines may not have found the necessary anchorage in culture. Perhaps, by having a mechanism that enables them to move, they

can explore the environment until they are in contact with enough surrounding adhesion molecules. It is also tempting to link such changes with synaptic plasticity. The molecular changes that allow the apparently random movements seen by Fischer *et al.* could be directly related to outgrowth of spinules in an excitatory synapse. The postsynaptic membrane of the excitatory synapses features a structure known as the postsynaptic density. This is a solid structure that sits directly under the postsynaptic membrane at the end of dendritic spines and contains the synaptic receptors as well as Ca^{2+} calmodulin-dependent kinase II, actin, tubulin and other structural molecules. It would presumably not be possible for an extending actin filament running up the length of the dendritic spine to grow beyond the density.

The postsynaptic density can take many forms when observed in three-dimensional reconstructions from electron micrographs. In small synapses it is a simple circle of electron-dense material, but it can have the shape of a horseshoe or various complex, larger versions. In some cases, spinules can be seen protruding through it. It has been suggested that these larger, 'perforated synapses' are more effective synapses, and that plastic change might result in the development from simple to perforated synapses⁴⁻⁶. Clearly, Fischer *et al.* have not observed changes in the size of the dendritic spines. However, their results indicate that the actin filaments that can be seen up the centre of the spine normally tend to grow and move continuously if physically unrestricted. In the brain, when they hit the well-tethered solid structure of the postsynaptic density, further growth would be physically limited and, possibly, controlled by actin-binding proteins. Only if space became available would they be able to grow further.

Constant polymerization and attempts towards growth in a flexible structure of limited size could result in exactly the writhing forms seen by Fischer *et al.*, if the dendritic spine were not fixed in place. Such an arrangement could conceivably be involved in synaptic plasticity. If, when plasticity is induced, an enzyme reaction occurs that allows a weakening of the postsynaptic density (maybe also affecting actin-binding proteins), actin filaments might immediately have some leeway for further growth and push the membrane out, forming a perforation through the density.

Thus, the dynamic changes in actin seen by Fischer *et al.* probably have physiological relevance, although whether it is to move the dendritic spines to the extent seen in culture is, I would suggest, rather doubtful. It is a fascinating observation, however, and the constant dynamics of actin within the dendritic spines could provide a mechanism

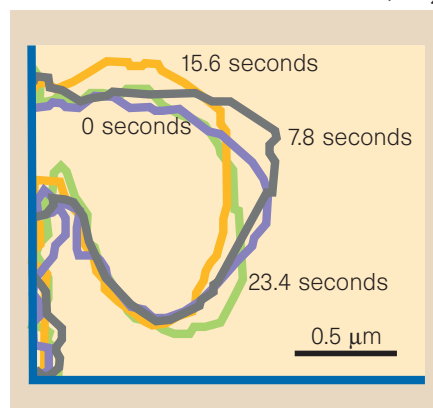


Figure 2 A sequence of spine movements, adapted from video imaging by Fischer *et al.*¹. Each coloured outline was traced from the outline of the same spine pictured over 23 seconds at the times indicated.

for the development of spinules or other morphological changes that may occur in synaptic plasticity. □

Frances A. Edwards is in the Department of Physiology, University College London, Gower Street, London WC1E 6BT, UK.
e-mail: f.a.edwards@ucl.ac.uk

1. Fischer, M., Kaech, S., Knutti, D. & Matus, A. *Neuron* **20**, 847–854 (1998).
2. Wagner, H.-J. *J. Neurocytol.* **9**, 573–590 (1980).
3. Ter-Margarian, A. & Djamgoz, M. B. A. *Neurosci. Lett.* **147**, 131–135 (1992).
4. Fifkova, E. *Cell. Mol. Neurobiol.* **5**, 47–63 (1985).
5. Geinisman, Y. *Hippocampus* **3**, 417–434 (1993).
6. Edwards, F. A. *Physiol. Rev.* **75**, 759–787 (1995).

Glacial cycles

Jive talking

John Chappell

Assessment of future climatic changes would be more secure if we fully understood the large or rapid climatic changes associated with the ice ages of the past 2½ million years. Because records of past climates become more sparse further back in time, the most closely studied period has been the last glacial period, which commenced about 115,000 years ago, oscillated, reached its climax of 20,000 years ago and ended 7,000–10,000 years ago.

The last ice age may not represent all modes of behaviour of the climate system, however, and one indication of more distant events is the amount of water locked into ice sheets at various times, which in turn is reflected in sea level. Hence the interest of the paper by Rohling *et al.* (page 162 of this issue¹) in which the authors have estimated sea-level lows, or lowstands, of the past 500,000 years. Most notably, they calculate that, during an ice age that peaked about 450,000 years ago (an interval known to stratigraphers as ‘stage 12’ or MIS12), sea level was around –140 m; that is, 140 m lower than it is today.

What that implies is that the quantity of ice at the climax of MIS12 was 15–20% greater than at the peak of the last ice age

(MIS2), when average sea level was about 120 m lower than now. The total ice mass on northern America and Europe in MIS2 was about 1.5 times greater than that covering Antarctica today; the low sea level in MIS12 implies a further quantity of ice which Rohling *et al.* equate with the volume of the present Greenland and West Antarctic ice sheets (about $5\text{--}7 \times 10^6 \text{ km}^3$). It is interesting to reflect that melting of these two ice sheets today would raise sea level to about +20 m, a level postulated for the interglacial stage (MIS11) which followed MIS12 (refs 2, 3; see Table 1 on facing page). Sea level at –140 m in MIS12 invites the question, ‘Where was the ice?’; and a level of +15 to +20 m in MIS11 asks, ‘What melted?’, with the corollary, ‘Could it happen now?’. Both events warrant close scrutiny.

Rohling *et al.* derive the –140-m sea level of MIS12 using the Red Sea as their gauge (Fig. 1). The Red Sea does not exchange water freely with the Gulf of Aden because of a sill, at a depth of 137 m, at the Strait of Bab-el-Mandab. So the salinity of the Red Sea is driven by evaporation and is critically affected by flow through the strait; salinity should increase if sea level falls. Sediment cores from the Red Sea’s floor contain fossil plankton

that indicate past salinity, but absences of fossils at certain levels, including MIS2 and MIS12, represent episodes when the sea water was too saline for planktonic life. Salt-balance calculations indicate that ‘aplanktonic’ events would occur if the Bab-el-Mandab sill was at about only 18 m; to achieve this, sea level would have to be about 120 m lower than today.

However, the sill appears to have been at greater depths in the past, because increasing proportions of microfauna that tolerate high salinity show that the Red Sea has become progressively more saline over the past 500,000 years. Rohling *et al.*¹ calculate that the Bab-el-Mandab strait was about 20 m deeper in MIS12; so the MIS12 aplanktonic event should represent a sea level of –20 m lower than that of the aplanktonic lowstand in MIS2, when sea level was about –120 m.

Oxygen-isotope data from microfossils in deep-sea cores lend support to this conclusion. The quantity of water locked into ice sheets affects the oxygen-isotopic composition of ‘average’ sea water, because ice sheets are depleted in the heavier isotope, ¹⁸O. Changes in the isotopic composition of sea water are recorded in microfossil carbonate, although temperature also affects the microfossil isotopes. Data from several cores⁴ indicate that melting of ice sheets after the last ice age (MIS2) caused the oxygen-isotope ratio of average sea water to decrease by 1.26‰, representing a sea-level rise from MIS2 to the present of 115–125 m; by comparison, the isotope ratio decreased by 1.44–1.51‰ through the MIS12–MIS11 transition, implying a sea-level rise of 130–150 m (assuming proportionality).

Evidence for past peaks in sea level comes from both the height of fossil reefs and oxygen-isotope data, but support is less strong for a sea-level high of +20 m in MIS11 (inferred from raised reefs at Sumba, Indonesia, and raised beaches elsewhere^{2,3}). Age measurements of the raised reefs and corrections for tectonic uplift have significant uncertainties. These uncertainties are likely to be compounded by isostatic movements, which are vertical earth movements caused by redistributing water from ocean to ice sheets, and have their largest effects on interglacial shorelines when the interglacial is long, as was MIS11, and when the preceding glaciation was large⁵. The oxygen-isotope data are ambiguous here: depletion of 0.1–0.4‰ relative to present values is observed for MIS11 in some (but not all) deep-sea cores^{4,6}, but this could equally reflect significantly higher sea level or a modest temperature increase of up to 1.5 °C.

Determination of sea level in MIS11 recalls Archimedes’ comment — ‘Give me a place to stand and I will weigh the Earth’. We may not yet have found the place to stand.



Figure 1 The Red Sea, which runs southeast to the ‘pinch point’ of the sill at the Strait of Bab-el-Mandab, before opening out into the Gulf of Aden. Rohling *et al.*¹ have used proxies for salinity in the Red Sea to estimate lows in sea level over the past 500,000 years.

Table 1 The relevant ice-age stages, at a glance

MIS12	Peak of an ice age 450,000 years ago. Sea level estimated by Rohling <i>et al.</i> ¹ at 140 m below that at present (–140 m).
MIS11	Interglacial period that followed MIS12, about 400,000 years ago. Sea-level estimate +20 m.
MIS2	Peak of the last ice age, 20,000 years ago. Sea-level estimate –120 m.

Nonetheless, the large rise in sea level after the MIS12 low and the long MIS11 interglacial both appear to differ from most other glacial–interglacial pairs, which is intriguing. The ice ages were strongly regulated by variations of the seasonal distribution of solar radiation (caused by precession of the Earth's axis and slow variations of both axial tilt and orbital eccentricity⁷). The strength of this pacemaker was relatively weak during MIS11 and at the climax of MIS12, and was also weak at the height of the last ice age (MIS2), when ice volume seems to have been 15–20% smaller than in MIS12.

The contrast suggests that the climate system may be more unpredictable when the pacemaker is weak than when it is strong². For all that we broadly know about the complex interactions between oceans, ice and climate⁷, climatic behaviour during the ice ages, seen in most detail for the last glacial cycle⁸, is more an imperfectly understood nervous jive than a stately ballet that

gracefully follows orbital forcing.

Rohling and colleagues' new data¹ from MIS12 tell us that our knowledge of the climate system is far from complete. The question 'Where was the ice?' compels further investigation. The possibility that sea level was up to 20 m higher in MIS11 urgently requires confirmation; if it is confirmed, current knowledge is challenged by the questions, 'What melted?' and 'Could it happen now?'

John Chappell is in the Research School of Earth Sciences, Australian National University, Canberra, ACT 0200, Australia.

e-mail: John.Chappell@anu.edu.au

1. Rohling, E. J. *et al.* *Nature* **394**, 162–165 (1998).
2. Pirazzoli, P. A. *Mar. Geol.* **109**, 221–236 (1993).
3. Howard, W. R. *Nature* **388**, 418–419 (1997).
4. Shackleton, N. J. *Quat. Sci. Rev.* **6**, 183–190 (1987).
5. Lambeck, K. & Nakada, M. *Nature* **357**, 125–128 (1992).
6. Oppo, D. W., Fairbanks, R. G., Gordon, A. L. & Shackleton, N. J. *Paleoceanography* **5**, 43–54 (1990).
7. Imbrie, J. *et al.* *Paleoceanography* **8**, 699–735 (1993).
8. Watts, W. A., Allen, J. & Huntley, B. *Quat. Sci. Rev.* **15**, 133–155 (1996).

Nanotechnology

A quantum leap for electronics

Lydia L. Sohn

Many recent advances in nanofabrication have fuelled the hope that electronic devices can be shrunk from the current micrometre-length scale all the way down to the single-atom or molecule scale (Fig 1). On page 154 of this issue¹, Scheer *et al.* describe electrical measurements on single metal atoms — the smallest electronic components possible. They show that the conductivity depends on the num-

ber of valence orbitals available in each atom.

By pushing devices to such an extreme, novel quantum phenomena — such as single-electron charging effects^{2,3}, or energy-level or conductance quantization^{2,3} — might readily be exploited near room temperature (at present, in devices orders of magnitude larger, such phenomena are manifest only at millikelvin temperatures). And molecule-based technology may yield

ultrafast, ultrasensitive hybridized devices based on materials other than traditional semiconductors or superconductors.

Already, a single C₆₀ molecule has been operated as an amplifier⁴, and a single carbon nanotube has been used to make a transistor that works at room temperature⁵. With a whole new class of electronic devices based on single atoms or molecules entirely within our technological reach, it is an exciting time for physics, engineering, material science, chemistry and even molecular biology. Gradually emerging from this excitement is a new field: molecular electronics.

But the future of molecular electronics depends on answers to a series of fundamental questions. First, what types of atom or molecule should one use, so that they can be integrated easily with the macroscopic world and robust enough to withstand millions of operations per second? (Should it be inorganic, organic or biologically based, for example?) Second, can we identify or develop effective technologies to manipulate these single atoms or molecules with precision, and integrate them into a macroscopic electronic circuit? For some time now, we have turned to the scanning tunnelling microscope (STM), which can manipulate single atoms or small groups of them, given atomically smooth substrates in ultra-high vacuum and at low temperature. The copper 'quantum corral'⁶ is one of the most powerful illustrations of this technology. But STMs are slow, and limited in terms of strict material and environmental requirements; there might be a better technology — faster, more encompassing, less restrictive.

Third, what are the macroscopic electron-transport properties of molecule-based devices? Finally, and most importantly, what particular molecular or atomic properties determine the electronic properties, such as conductance?

For single metal atoms, Scheer *et al.*¹ give answers to the last two questions. Mainly by

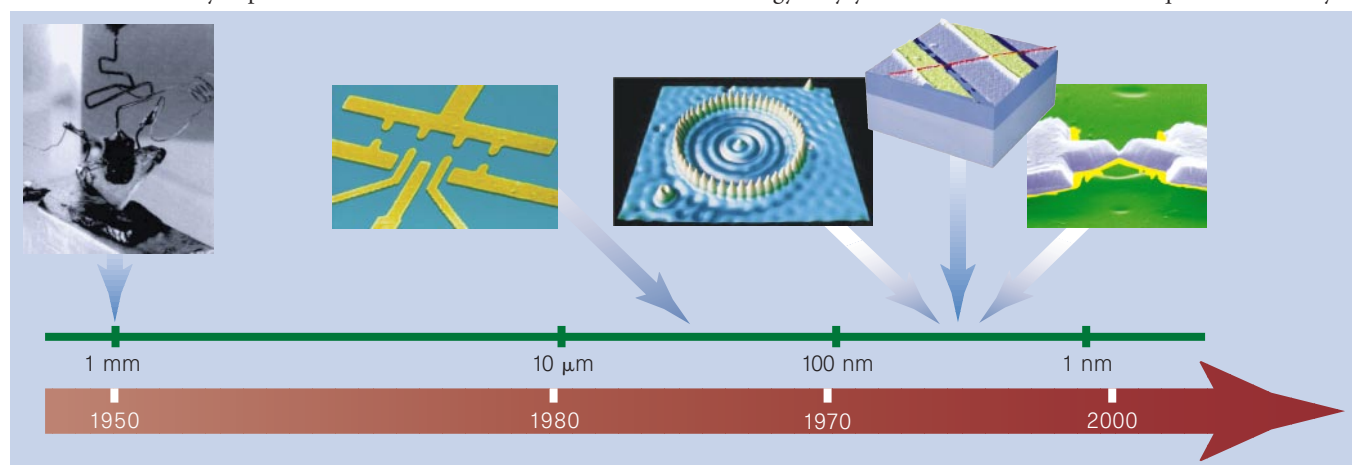


Figure 1 The shrinkage of electronic components. The length scale reached by technology has dropped steadily from the millimetre scale of the early 1950s to the present-day atomic scale. The representative devices, from left

to right, are: the first transistor, a quantum-dot turnstile, a copper 'quantum corral', a carbon-nanotube transistor, and the latest — a one-atom point contact.

using mechanically controllable break junctions, they were able to make and measure the simplest type of single-atom device: a one-atom constriction between two metallic leads. By measuring the electronic properties of the devices as a function of distance (this is accomplished by stretching each device to breaking), the authors determined that the maximum number of observable conduction channels N in a one-atom contact is equal to the number of valence orbitals N_{orb} for the particular atom. For example, they showed that lead, whose valence shell contains 6s and 6p orbitals ($N_{\text{orb}} = 4$), has four conducting channels, whereas gold, with only a 6s orbital ($N_{\text{orb}} = 1$), has one channel. Slight deviations from the general rule $N = N_{\text{orb}}$ can be explained by taking into account the atomic orbital structure coupled with the local geometry of the single-atom contacts.

The impact of Scheer *et al.*'s experiments is threefold. First and foremost, they show that single-atom devices can be made by lithography. Second, they provide the first experimental verification that local electronic properties of single atoms or molecules strongly affect conduction through

macroscopic leads. Third, they provide a precedent for using such 'spectroscopic' techniques to sense and identify single molecules for applications in other fields, such as biotechnology and chemical synthesis (L. L. Sohn & J. B. Knight, in preparation).

The future success of molecular electronics relies on a collaboration between disparate disciplines. Already, a number of research groups in vastly different areas are working together to cultivate this new field. Will molecular electronics be the electronics revolution of the twenty-first century? Or will it merely be a passing fad or academic exercise? It is too early to tell definitively, but the early results look promising. □

Lydia L. Sohn is in the Department of Physics, Princeton University, Princeton, New Jersey 08544, USA.

e-mail: sohn@oberon.princeton.edu

1. Scheer, E. *et al.* *Nature* **394**, 154–157 (1998).
2. Grabert, H. & Devoret, M. H. (eds) *Single Charge Tunneling* (Plenum, New York, 1992).
3. Sohn, L. L., Kouwenhoven, L. P. & Schön, G. (eds) *Mesoscopic Electron Transport* (Kluwer, Boston, 1997).
4. Joachim, C. & Gimzewski, J. M. *Chem. Phys. Lett.* **265**, 353–357 (1997).
5. Tans, S. J. *et al.* *Nature* **386**, 474–477 (1997).
6. Crommie, M. F., Lutz, C. P. & Eigler, D. M. *Science* **262**, 218–220 (1993).

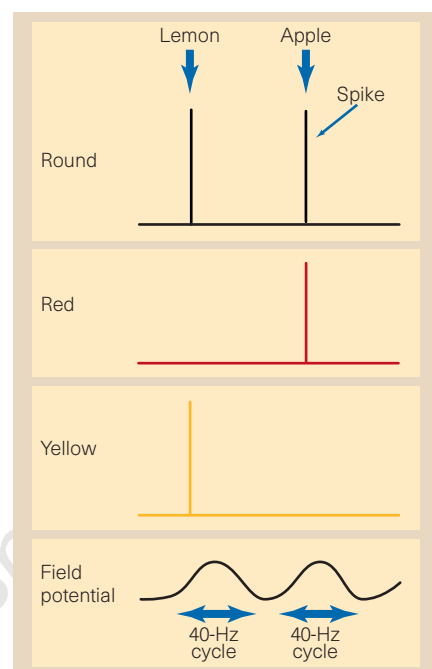


Figure 1 Binding and segmentation of 40-Hz oscillations. An object is defined by several features. The binding of these features for a single object occurs because the neurons that detect these features (round, red and yellow, for example) fire spikes in the same 40-Hz cycle. Segmentation means that the spikes that signify different objects fire at different times, and here that means different 40-Hz cycles. 40-Hz oscillations are detected in field recordings that reflect the synchronized firing of groups of neurons clocked by 40 Hz. Note that individual neurons do not necessarily fire at 40 Hz.

Neuroscience

What makes the brain's tickers tick

John Lisman

Seventy-five years ago, Berger placed electrodes on his son's scalp and made the first observation of brain waves. The frequency of these waves was ten cycles per second (10 Hz), but since then oscillations at other frequencies (5, 40 and 200 Hz) have been detected. These oscillations involve the synchronized firing of large groups of neurons, suggesting a role in clocking, yet remarkably little is known about their function or mechanism. Oscillations do not normally occur in the brain slices that are routinely used to investigate mechanism. But recent studies — including reports by Fisahn *et al.*¹ and Draguhn *et al.*² on pages 186 and 189 of this issue — have succeeded in inducing 40- and 200-Hz oscillations, respectively, in hippocampal slices. These studies provide new insights into the mechanisms of oscillations, and these have interesting implications for function.

Fisahn *et al.*¹ provide the strongest evidence to date that 40-Hz oscillations are due to a feedback-inhibition mechanism³ — the firing of pyramidal cells excites inhibitory interneurons, which then inhibit the pyramidal cells. After inhibition decays, the pyramidal cells fire again, initiating the next cycle. The key new evidence is that oscillations are prevented if the fast excitation of interneurons by pyramidal cells is blocked. This observation was made in the CA3 region

of the hippocampal slice, but an entirely different result has been obtained by others in the CA1 region⁴, where the oscillation persists when fast excitation is prevented. This persisting oscillation is due to an interneuron network in which a steady depolarization causes interneurons to fire and mutually inhibit one another. The next cycle of firing then occurs after this inhibition decays^{4,5}. *In vivo* studies should tell us whether either or both of these mechanisms operate under particular physiological conditions.

The most widely accepted hypothesis about the function of 40-Hz oscillations is that they reflect the synchronization of cell firing that 'binds' together the elementary features of an object^{6,7}. Consider a simplified example in which there are three neuronal feature-detectors that respond to round, red and yellow, respectively (Fig. 1). When a lemon is present, the yellow and round cells fire synchronously. By detecting this synchrony, downstream neurons could 'know' that a lemon is present. Oscillations may also 'segment' information if more than one object is present. For example, if both an apple and a lemon are present, the cells representing these objects fire at different phases.

But what exactly is meant by phase, and on what timescale is synchrony defined? One possibility is that synchrony means firing together within a period of a few milliseconds,

and phase is defined by the part of a 40-Hz cycle during which firing occurs. Alternatively, synchrony may be defined as firing at any time during a given 40-Hz cycle, and phase may be defined by the particular 40-Hz cycle during which firing occurs. Several lines of information support this latter view. First, if the brief interval between two clicks is gradually increased, the sound is initially perceived as a single click, but then as two. This transition coincides with the appearance of a second 40-Hz cycle⁸. Second, recordings in the hippocampus and associated cortical regions show that roughly seven cycles of 40 Hz are nested within the period of a slower, 5-Hz, 'theta' oscillation⁹, generating a limited number of discrete phases (about seven) for 'items' (in this region, the 'items' are memories). This limit may explain the limit of 7 ± 2 items that can be held active in short-term memory¹⁰. Moreover, if different memories are active in different 40-Hz cycles, scanning through each additional item held in short-term memory should take an extra amount of time that is comparable to the 40-Hz period¹⁰ — and this is what is psychophysically measured. Finally, quantitative properties of hippocampal place cells can be explained if the recall of sequential expected locations

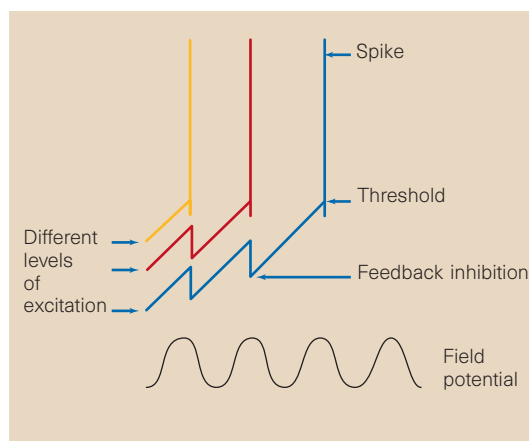


Figure 2 Mechanism by which a negative-feedback version of 40-Hz oscillation could serially activate memories in sequential 40-Hz cycles. Each trace represents the membrane potential in a cell that encodes a given memory. After the most excited cell fires, it causes feedback inhibition (down phase of intracellular 40 Hz) and cannot be excited again (trace no longer shown). As the inhibition wears off, the next most excited cell reaches threshold. A memory is represented by a group of neurons whose synchronized firing generates the field potential.

occurs in successive 40-Hz cycles¹¹. These results, and related studies¹² in invertebrates, indicate that 40-Hz oscillations separate different units of information and organize the serial activation of these units.

Although such serial activation brings to mind the operation of a digital computer, simple models have been developed for how this could be done by neural networks. The models are based on the 'winner-take-all' process¹³ that is implicit¹⁰ in the negative-feedback model of 40-Hz generation³. Memories are assumed to have different levels of excitatory drive — the problem is how to get them to fire in separate 40-Hz cycles, and Fig. 2 shows how this could be achieved. During the inhibitory phase of oscillations, all neurons are shut off. After inhibition partially decays, the most excitable group (the 'winner') reaches threshold first, inhibits all of the others by feedback inhibition and then inactivates itself (owing to a hyperpolarizing after-potential, for instance). On the next cycle, the next most excitable group will fire and, in this way, memories can be serially read out.

Inhibition is vital for all models of 40-Hz oscillation, but it is unlikely that inhibition can account for the much faster 200-Hz oscillations^{3,14}. These oscillations occur during sharp waves¹⁵ — events that may be an ultra-fast form of memory recall, important for transferring information from the hippocampus to the cortex. Draguhn *et al.*² provide the first information about the mechanism of 200-Hz oscillations, concluding that synchronization is due to gap junctions between nearby axons. This makes sense because the direct flow of current between cells that occurs at gap junctions is much faster than chemical synaptic transmission, so it is appropriate for rapid synchronization. What remains unclear is the function of this synchronization. Without it, two cells firing at 200 Hz would differ in firing time by no more than 2–3 milliseconds. Downstream neurons are not thought to be affected by jitter this small, but perhaps this assumption is wrong. What is clear is that this and other questions about brain oscillations can now be approached using brain

slices, a development that will help us to understand what makes the brain tick. □

John Lisman is in the Department of Biology, Brandeis University, Waltham, Massachusetts 02254, USA.

e-mail: lisman@binah.cc.brandeis.edu

1. Fisahn, A., Pike, F. G., Buhl, E. H. & Paulsen, O. *Nature* **394**, 186–189 (1998).
2. Draguhn, A., Traub, R. D., Schmitz, D. & Jefferys, J. G. R. *Nature* **394**, 189–192 (1998).
3. Freeman, W. J. *J. Neurophysiol.* **31**, 337–348 (1968).
4. Whittington, M. A., Traub, R. D. & Jefferys, J. G. R. *Nature* **373**, 612–615 (1995).
5. Wang, X.-J. & Buzsáki, G. *J. Neurosci.* **16**, 6402–6413 (1996).
6. Singer, W. & Gray, C. M. *Annu. Rev. Neurosci.* **18**, 555–586 (1995).
7. von der Malsberg, C. *Curr. Opin. Neurobiol.* **5**, 520–526 (1995).
8. Joliot, M., Ribery, R. & Llinas, R. *Proc. Natl Acad. Sci. USA* **91**, 11748–11751 (1994).
9. Bragin, A. *et al.* *J. Neurosci.* **15**, 47–60 (1995).
10. Lisman, J. E. & Idiart, M. A. *Science* **267**, 1512–1515 (1995).
11. Jensen, O. & Lisman, J. E. *Learning and Memory* **3**, 264–278, 279–287 (1996).
12. MacLeod, K. & Laurent, G. *Science* **274**, 976–979 (1996).
13. Coultop, R., Granger, R. & Lynch, G. *Neural Networks* **5**, 47–54 (1992).
14. Chow, C., Ritt, J., Sott-Trevino, C. & Kopell, N. *J. Computational Neurosci.* **5**, 5–16 (1998).
15. Chrobak, J. J. & Buzsáki, G. *J. Neurosci.* **16**, 3056–3066 (1996).

Retraction

I wish to point out that I no longer stand by the views reported in my News and Views article "Immunology: Ways around rejection" (*Nature* **377**, 576–577; 1995), which dealt with a paper in the same issue ("A role for CD95 ligand in preventing graft rejection" by D. Bellgrau *et al.* – *Nature* **377**, 630–635; 1995). My colleagues and I have been unable to reproduce some of the results of Bellgrau *et al.*, as reported by J. Allison *et al.* (*Proc. Natl Acad. Sci. USA* **94**, 3943–3947; 1997). □

David L. Vaux, The Walter and Eliza Hall Institute of Medical Research, Post Office RMH, Victoria 3050, Australia.
e-mail: vaux@wehi.edu.au

D. Bellgrau *et al.* (e-mail: don.bellgrau@uchsc.edu) consider that their results are reproducible and stand by them. They note, however, that the magnification in Figure 1g of their paper should be 113 times, not 45 times as printed. Both groups believe that other published data support their views, and interested readers can contact them directly for further details. —Editor, *Nature*.

Daedalus

Corrugated carbon

Last week Daedalus was devising a method of making carbon nanotubes to order, by the electrolytic deposition of C₂ radicals. His subtle catalytic anode had circular reaction sites as templates, to assemble the radicals into nanotubes. These would grow out from the anode as a close-packed 'fur' of parallel nanotubes, which could be stripped off and made into threads and fabrics like any other staple fibre.

He is now generalizing the idea. Already carbon membrane structures can be fabricated by deposition on an alumina template; an anodic template could do so far more controllably. It would assemble whatever carbon structure was defined by the pattern of catalytic sites laid down on it. In particular, if it bore a hexagonal lattice of such sites, it could assemble a carbon honeycomb. Imagine, says Daedalus, a plane hexagonal lattice 'ground plan'; and imagine each line on this lattice extruding a graphite monolayer sheet upwards out of the plane. Each sheet will touch two others at 120°. If the carbon atoms forming the junctions are bonded together by the tetrahedral sigma bonds found in diamond, the result will be carbon honeycomb — a giant polymeric generalization of the triptycene molecule.

The holes running through a block of carbon honeycomb will have the size defined by the hexagonal catalytic pattern laid down on the anode that forms it. Even the best modern microfabrication methods would produce a catalytic array very coarse on the molecular scale. Finer arrays might be made by cunning surface-crystallization methods, or painstaking assembly by atomic-force microscope. They might even be given two different pore sizes, alternating across the lattice.

So carbon honeycomb could be given a wide range of useful properties. With the finest pores, it would be structurally superb: combining the stiffness of diamond with the oriented strength and toughness of carbon fibre. Larger pores would allow ions and molecules to traverse them, giving novel one-dimensional ionic conductors for batteries, shape-selective catalysts, and molecular sieves and absorbents of vast capacity. The coarser grades would be increasingly tenuous but still very strong, a sort of oriented carbon aerogel. They would be ideal for the insulating interior of laminated panelling and light-weight aerospace components. Filled with epoxy or polyester 'honey' and set solid, they would give an outstanding reinforced composite.

David Jones

Obituary

Alberto P. Calderón (1920–98)

Mathematician and master of singular integrals

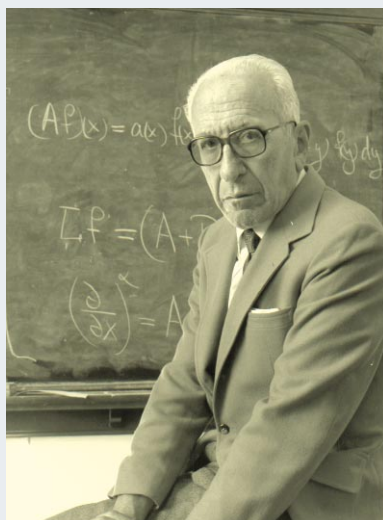
Alberto Calderón, one of the twentieth century's greatest mathematicians, died on 16 April this year after a brief illness. He was 77. Calderón was a pioneer in the areas of Fourier analysis and partial differential equations, and their interconnections.

Born in 1920, in Mendoza, Argentina, Calderón received his early education in Switzerland as well as in his native country. He began his career as an engineer, and one can clearly see the resulting influence in his mathematics and in the huge impact of the applications of his work.

Calderón's transition from engineer to mathematician is associated with a wonderful story, involving Antoni Zygmund, his future mentor. In 1948, Zygmund — a giant of mathematical analysis, and author of the leading book on Fourier analysis — delivered some lectures in Buenos Aires. During one of the lectures, attended by Calderón, Zygmund reproduced the proof of one of the most fundamental results of the field, the boundedness of the Hilbert transform on L^p . The Hilbert transform is one of the most important operators of analysis. At the same time, it expresses the passage from the real to the imaginary part of a complex analytic function, and is intimately related to the partial sums of a Fourier series. L^p estimates are estimates of the size of functions in terms of the integral of the p th power of their absolute value.

Calderón approached Zygmund after the lecture, asking why the argument given in the talk was so much more complicated than that in the book. Zygmund, who was completely mystified, replied that the proof presented in the talk *was* the one in the book. It turned out that Calderón had studied Zygmund's book according to his usual custom, which was to cover up the proof of each theorem, trying to produce his own argument, and then checking back to compare it with the book. In this case, Calderón had forgotten to check the proof in Zygmund's book, and had come up with a far superior argument. When it became clear to Zygmund what had happened, he recognized Calderón's potential, and persuaded him to study mathematics at the University of Chicago, where Zygmund was professor.

This was the start of one of the most fruitful and celebrated collaborations in the history of mathematics. Calderón



earned his PhD under Zygmund's guidance in 1950, and they soon embarked on work that was to revolutionize analysis — the theory (now known as the Calderón–Zygmund theory) of singular integrals. It turns out that many of the most important processes of nature (such as electrical and heat conduction) are expressed in terms of certain integrals which look divergent or infinite. However, when properly interpreted, these integrals possess just the right cancellation of their positive and negative parts so that they are actually not merely well defined and finite, but well behaved.

These singular integrals have as their simplest example (in one dimension) the Hilbert transform. The proof that the Hilbert transform is well behaved was initially accomplished by arguments involving the theory of analytic functions of a complex variable. However, for a great many applications, one needs to handle higher-dimensional singular integrals, where complex methods do not apply. Calderón and Zygmund developed an entire programme, based on real variable theory, to analyse singular integrals, and this spectacular accomplishment came to fruition in 1952 with the appearance of a classic paper in *Acta Mathematica*, “On the existence of certain singular integrals”.

Applications of tremendous significance appeared immediately. For example, the analysis (in terms of L^p estimates) of the most important partial differential operator, the Laplacian, was made possible by this theory. Later, Calderón observed that the Littlewood–Paley theory, one of the most sophisticated tools of harmonic analysis at that time, was a simple consequence of the theory of singular integrals, if only one was

willing to view the integrals as being vector valued. Another remarkable application came in the form of pseudodifferential operators and their use in geometry, most notably in addressing the Atiyah–Singer index theorem. Calderón was also able to apply singular integrals to prove the uniqueness of a solution for a very general class of partial differential equations with respect to the so-called Cauchy problem.

There were numerous further applications, too many to mention here. But one general remark should be made about the proof of the Calderón–Zygmund theorem. The proof proceeds by showing how to take a function, and cut it up into its large and small parts, a method known as the Calderón–Zygmund decomposition. This sounds simple, but it turns out that the correct way to split a function is not at all obvious. One needs to come to grips with the averages of a function over cubes, rather than just its value at individual points. This understanding of the splitting of functions, obtained by Calderón and Zygmund, is even more basic than any individual application.

Calderón had a marked tendency to be the exception to any given rule. For instance, mathematics is said to be a young person's game, but some of Calderón's greatest work was done when he was nearly 60 years old. In an attempt to solve elliptic boundary value problems in general, and to understand the theory of complex analytic functions in domains with sharp corners, he courageously attacked problems involving singular integrals which were not translation invariant. His brilliant success in the late 1970s represented proof positive that he was still at the height of his powers at an age when many people are contemplating retirement.

Calderón the man was as exceptional as Calderón the mathematician. He was extremely distinguished looking, and could seem aloof and distant to students and even some colleagues. But to those who knew him at all, he was remarkable for his warmth and kindness. He had a keen sense of humour, and was strikingly modest; above all, he had the greatest inner strength and character. This wonderful man, who touched so many by his mathematical work, will be remembered and sorely missed by those who had his friendship and knew him by his generosity.

Robert A. Fefferman

Robert A. Fefferman is in the Department of Mathematics, University of Chicago, 5734 South University Avenue, Chicago, Illinois 60637-1546, USA.

e-mail: raf@math.uchicago.edu

Lane's landscapes

Biologists such as Nancy Lane are venturing into previously unexplored and strangely beautiful realms of the cell, using sophisticated microscopes allied with familiar, age-old visual techniques.

Martin Kemp

Just as Robert Hooke in the seventeenth century was beguiled by the microcosmic wonders of the world in miniature disclosed by his microscope, so the successively refined techniques of modern electron microscopy have repeatedly occasioned visual excitement in pioneers and practitioners alike. And, in keeping with the perceptual practices of those who first looked through optical devices, the topographies of the unfamiliar worlds are certified by analogy to morphological features visible within the normal compass of our unaided vision. The metaphorical language of science often speaks of this mode of seeing and describing.

When Gerd Binnig viewed the multiple diamond pattern of the 7×7 surface of silicon in the scanning tunnelling microscope, which he had invented with Heinrich Rohrer in 1982, he responded rhapsodically, talking in terms of aesthetic revelation: "Here one saw little hills, and the hills formed a complicated pattern with this symmetry. We like symmetry. If something is symmetrical, it is very pretty. This was complicated but regular, extremely surprising and pretty."

The less overtly symmetrical landscapes of organic tissues are equally 'pretty' in their own way. The Cambridge cell biologist, Nancy Lane, talks of the freeze-fracture replicas of plasma membranes as "providing stunning evidence, if such were required, of the beauty of the cell". Yet it is a beauty that can only be disclosed by complex techniques of preparation.

Lane's technique involves the fast freezing of plasma membranes of eukaryotic cells from invertebrates, mainly arthropods, which have been treated with 'antifreeze' to



Lane's "Detail showing septate junctional intramembranous particles in linear arrays" ($\times 65,000$).



Lane's "Freeze-fracture replica of plasma membranes in a septate junctional area from the midgut of the moth, *Manduca sexta*" ($\times 28,000$).

avoid crystal formation. The two adhering membranes that make up a junction are cleaved with a knife in an evacuated chamber, the fracture plane running medially through either membrane. The resulting relief surfaces, each of which may carry some residues from the other half of the membrane, are then shadowed with a heavy metal 'sprayed' at an angle onto the sides of the hills and valleys. The surface is next 'faced up' with a thin film of carbon — non-opaque to electrons — and the biological tissue is removed when the preparation has been thawed. The metallic "death mask", as Lane calls it, is then examined under a transmission electron microscope.

What we see in the microdomains are the configurations of intramembranous particles that form the intercellular junctions. In some places we see the prominences of the proteins anchored in their 'sea' of lipid. In others we see the depressions which retain, in negative, the imprint of the projecting fea-

tures from the half-membrane that has been cleaved away.

Reading the complex and subtle configurations of convexities and hollows involves considerable visual agility, accentuating the positive and, unlike the popular song, not eliminating but asserting the negative.

Navigating through the landscape, devoid of the recognizable features of our human terrain, requires disciplined scrutiny, the cultivation of visual memory and considerable perceptual control — and the 'correct' orientation of the highlights and metallic shadows as if lit, ideally, from the upper left, if prominence and hollow are not to be misleadingly reversed. Newly accessible territories are yielding their secrets through the exercise of enduring visual habits. □

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.

Nicotine withdrawal and accident rates

It is well known that when regular smokers quit smoking, their mood and cognitive performance typically deteriorate within a few hours of abstaining^{1,2}. But do these psychological deficits, readily measured in the laboratory, cause major disruption in everyday activities, such as performance at work? Here we use a new method to address this question; our data suggest that the effects of nicotine withdrawal can indeed be detected in daily activities.

Since 1984, the second Wednesday of March has been known as No Smoking Day (NSD) in the United Kingdom. There is evidence^{3,4} that a large number of smokers (up to 2 million) make an attempt to abstain from smoking, or smoke less, on this day. We therefore decided to test the hypothesis that nicotine withdrawal causes deficits in real-world psychomotor performance by comparing a measure of performance on NSD itself with performance on Wednesdays before and after NSD.

Our measure was the number of non-fatal accidents at work reported to the Health and Safety Executive in Great Britain in industries in which the executive is the enforcing authority. Workplace accidents resulting in a major injury, or in a person having to take more than three days off work, have, by law, to be reported to this body in Great Britain. The national Health and Safety Executive statistics are based on reports under the Reporting of Injuries, Diseases and Dangerous Occurrences Regulations⁵.

Our assumptions were as follows: first, there will be many more people in the working population suffering from nicotine withdrawal on NSD itself than on Wednesdays before and after; and second, deterioration in psychological function as a result of nicotine withdrawal causes an increased chance of an accident at work.

Table 1 Numbers of reported non-fatal accidents in 1987–96

Day	'Before' NSD week	NSD week	'After' NSD week
Monday	695.9 (28.2)	690.3 (28.5)	646.6 (24.1)
Tuesday	593.3 (22.6)	586.1 (19.8)	555.4 (19.9)
Wednesday	553.4 (17.6)	575.8 (16.7)	519.1 (18.2)
Thursday	546.8 (16.7)	529.4 (20.9)	493.5 (19.3)
Friday	493.4 (16.8)	490.5 (23.4)	443.5* (20.1)

*We excluded data from Fridays that were UK Bank Holidays when computing this Friday value (the second Friday after NSD week in 1989 and 1991 was a 'Good Friday'). Bold text distinguishes Wednesdays before and after NSD week (bold) and NSD itself (bold italics). Data shown are means $\pm$ s.e.m.

We compared the average number of reported accidents occurring during NSD week with the average for the two weeks before NSD week (our 'before' measure) and the average for the two weeks after NSD week (our 'after' measure). Table 1 shows the means and standard errors of these measures for the ten years (1987–96) for which daily data were available. Planned paired-sample Student's *t*-tests (one-tailed) indicated that there were significantly more reported accidents on NSD than on the Wednesday in the 'before' measure ($t(9) = -2.15$, $P=0.03$), or on the Wednesday in the 'after' measure ($t(9) = -5.95$, $P<0.001$).

Visual inspection of data in Table 1 shows that, in contrast with NSD, there were actually fewer reported accidents on the Monday, Tuesday, Thursday and Friday in NSD week than on those days in the 'before' measure. This indicates that there was nothing about the week of NSD generally that was generating an excess of reported accidents. Moreover, although there are fewer reported accidents generally in the 'after' measure (perhaps because of better weather or more people being on holiday as spring progresses), the difference in the number of reported accidents on NSD compared with on the Wednesday in the 'after' measure is larger than for all the other days. Finally, there are also large day-of-the-week effects, for which there may be many underlying causes.

Of course, there may be some other factor associated with NSD that underlies the

effect seen. It is difficult, however, to imagine what it might be, especially given that the actual date of NSD can vary between 8 and 14 March across years. We stress that the NSD effect, if real, should not be construed as indicating that cessation attempts on No Smoking Day are a bad idea, although it may suggest that wider use of nicotine replacement might be beneficial. Our findings could be tested with data from other countries (for example, from the Great American Smokeout) and by examining other variables for which accurate day-by-day data are available (such as road traffic accidents).

Andrew J. Waters

Tobacco Research Section,
National Addiction Centre, Institute of Psychiatry,
London SE5 8AF, UK
e mail: rmjdajw@ucl.ac.uk

Martin J. Jarvis, Stephen R. Sutton

ICRF Health Behaviour Unit,
Department of Epidemiology and Public Health,
University College London, London WC1E 6BT, UK

1. Snyder, F. R., Davis, F. C. & Henningfield, J. E. *Drug Alcohol Dep.* **23**, 259–266 (1989).
2. West, R. J., Jarvis, M. J., Russell, M. A. H., Carruthers, M. E. & Feyerabend, C. *Br. J. Addict.* **79**, 215–219 (1984).
3. McGuire, C. *Pausing for Breath: A Review of No Smoking Day Research 1984 to 1991* (Health Education Authority, London, 1992).
4. West, R. J. *Getting Serious about Stopping Smoking: A Review of Products, Services, and Techniques. A Report for No Smoking Day* (Health Education Authority, London, 1997).
5. *A Guide to the Reporting of Injuries, Diseases, and Dangerous Occurrences Regulations 1985* (Health and Safety Executive, HMSO, London, 1986).

Cause of sea fan death in the West Indies

A fungus from the genus *Aspergillus* is the probable agent of epizootic infections that have caused mass mortality of sea fan corals (*Gorgonia ventalina*) over the past 15 years^{1,2}. Here we show that four strains of the fungus involved in these infections are members of the species *Aspergillus sydowii*, a common saprobe (an organism that lives on decaying matter) that is found in both terrestrial and marine environments. Isolates of *A. sydowii* taken from diseased sea fans caused new infections of sea fans in

inoculation experiments, whereas isolates taken from elsewhere did not.

We took four fungal isolates from diseased sea fans in Saba, Trinidad, the Florida Keys, and San Salvador, Bahamas. We cultured the isolates on agar media and analysed them morphologically and by DNA sequencing. The sequence of the 18S nuclear ribosomal small subunit gene was identical in three isolates tested, and was more similar to that of *Aspergillus fumigatus* than to any other sequence in GenBank³. All four isolates showed morphological characteristics typical of *A. sydowii*^{3,4}, including the production of conspicuously echinulate asexual spores (conidia) and a distinctive turquoise-blue colony colour (Fig. 1).

To establish that these isolates were *A. sydowii*, we sequenced a ~485-base-pair portion of the 5' non-translated region of the *trpC* gene from three of the isolates that could reinfect sea fans (see below). We compared these sequences to those from known *A. sydowii* isolates from peanuts in Georgia (isolate H640), dates in Austria (NRRL 242), dried fish in Japan (NRRL 244) and silk thread in Philadelphia (NRRL 249) and to sequences from other closely related species (*Aspergillus versicolor*, *Emericella nidulans* and *Emericella violacea*).

We chose this sequence because it shows 3.8–5.4% variation between phenotypic sister species of *Aspergillus* (*A. parasiticus* and *A. flavus*) and a maximum of 2.2% varia-



Figure 1 Colonial growth of the Saba sea fan pathogenic isolate in culture.

tion among isolates within a phenotypic species (*A. flavus*)⁵. *A. versicolor* was chosen because it is phenotypically nearly indistinguishable from *A. sydowii*, and *E. nidulans* and *E. violacea* were chosen as outgroups.

The sequences were subjected to maximum parsimony analysis, and all *A. sydowii* isolates and sea fan pathogenic isolates formed a strongly supported (100% bootstrap) clade that was separate from *A. versicolor* and the *Emericella* outgroups, with no evidence for distinction between the known pathogenic isolates and *A. sydowii* isolates from non-marine sources (Fig. 2). Sequence divergence between the sea fan pathogenic and non-marine isolates showed a maximum of only 1.1% (representing six variable nucleotide sites), whereas it was a minimum of 10.1% between any of these isolates and *A. versicolor* isolate 226. We conclude that the fungus causing mortality of West Indian sea fans is *A. sydowii*.

A. sydowii is a common, cosmopolitan saprobic fungus not previously recognized to cause widespread disease in plants or animals. It has been isolated from a variety of terrestrial environments, including deglaciated soils in Alaska and various soils in tropical regions, as well as from many different plant materials⁶. In a survey of fungi that could be cultured from subtropical marine waters near the Bahamas and the Straits of

Florida, *A. sydowii* was found in both eu-littoral zones and oceanic zones, and was isolated from waters collected as deep as 4,450 m (ref. 7). Many *Aspergillus* species, including *A. sydowii*, are tolerant of salt concentrations as high as, or higher than, those found in the marine environment.

To further test the virulence of *A. sydowii*, we used isolates from non-marine environments (NRRL 242, 244 and 249) and two isolates from sea fans (those from Saba and San Salvador) to challenge small (~20 cm in height) *G. ventalina* colonies in flow-through seawater aquaria at the Bahamian Field Station in San Salvador. We directly inoculated sea fan colonies with pure *A. sydowii* cultures in separate experimental aquaria and inoculated colonies with sterile media in control aquaria. All replicates inoculated with the Saba and San Salvador isolates showed recession of coenenchyme at the point of inoculation, exposing the axial gorgonin skeleton (central core). These are typical signs of aspergillosis of sea fans². This response did not occur in colonies inoculated with non-marine strains of *A. sydowii* or in controls. We conclude that, under these conditions, isolates of *A. sydowii* taken from diseased sea fans possess pathogenic potential not seen in isolates taken from other sources.

Several *Aspergillus* species have a tendency to be opportunistic animal pathogens, generally striking individuals with compromised immune systems. The infection of West Indian sea fans by *A. sydowii* may also be opportunistic, owing to weakening of the host from stress imposed by pollution, other pathogens or environmental factors. The fact that non-marine isolates of *A. sydowii* failed to cause disease may reflect the presence of a specific pathogenicity factor in some isolates, perhaps a mycotoxin or other secondary metabolite. The virulence of *A. sydowii* isolates taken from the water column should be tested to determine whether pathogenic potential exists in all marine isolates. It should also be determined whether isolates of other *Aspergillus* species, many of

which are commonly found in marine environments, are able to cause diseases of sea fans or other corals.

David M. Geiser*, John W. Taylor

Department of Plant and Microbial Biology,
University of California, Berkeley,
California 94720, USA

e-mail: jttaylor@socrates.berkeley.edu

* Present address: Department of Plant Pathology,

204 Buckhout Laboratory,
The Pennsylvania State University,
University Park, Pennsylvania 16802, USA
e-mail: dgeiser@psu.edu

Kim B. Ritchie

Biology Department, University of North Carolina,
Chapel Hill, North Carolina 27599, USA

e-mail: kritchier@mailserv01.isis.unc.edu

Garriet W. Smith

Department of Biology, University of South
Carolina-Aiken, Aiken, South Carolina 29801, USA
e-mail: garries@usca.usca.sc.edu

1. Nagelkerken, I. et al. in *Proc. 8th Int. Coral Reef Symp.* Vol. 1 (eds Lessios, H. A. & Macintyre, I. G.) 679–682 (STRI, Panama City, 1997).
2. Smith, G. et al. *Nature* **383**, 487 (1996).
3. Klich, M. A. & Pitt, J. I. *A Laboratory Guide to Common Aspergillus Species and their Teleomorphs* (CSIRO Division of Food Processing, North Ryde, Australia, 1988).
4. Raper, K. B. & Fennell, D. I. *The Genus Aspergillus* (Williams & Wilkins, Baltimore, 1965).
5. Geiser, D. M., Pitt, J. I. & Taylor, J. W. *Proc. Natl Acad. Sci. USA* **95**, 388–393 (1998).
6. Domsch, K. H., Gams, W. & Anderson, T.-H. *Compendium of the Soil Fungi* Vol. 1 (Academic, London, 1980).
7. Roth, F. J. et al. *Can. J. Bot.* **42**, 375–383 (1964).

Origins of Old Testament priests

According to Jewish tradition, following the Exodus from Egypt, males of the tribe of Levi, of which Moses was a member, were assigned special religious responsibilities, and male descendants of Aaron, his brother, were selected to serve as Priests (Cohanim). To the extent that patrilineal inheritance has been followed since sometime around the Temple period (roughly 3,000–2,000 years before present), Y chromosomes of present-day Cohanim and Levites should not only be distinguishable from those of other Jews¹, but — given the dispersion of the priesthood following the Temple's destruction — they should derive from a common ancestral type no more recently than the Temple period. Here we show that although Levite Y chromosomes are diverse, Cohen chromosomes are homogeneous. We trace the origin of Cohen chromosomes to about 3,000 years before present, early during the Temple period.

We characterized Y-chromosome-specific variation at six microsatellites (repeats of short nucleotide sequences) and six presumably 'unique-event' polymorphisms (UEPs) in a sample of 306 male Jews

Figure 2 Parsimony phylogenetic analysis of sequences from the 5' non-translated region of the *trpC* gene. The DNA region was amplified using the polymerase chain reaction and sequenced with the primers *trpC1* (5'-GACGGGAAATAG-GCTTCC) and *trpC3* (5'-CGC-CTTGTTGGGATGGTG) as described⁵. The aligned sequences were subjected to parsimony analysis using the PAUP v3.11 software package (D. Swofford). One of the two equally most parsimonious trees produced (length 308 steps) is shown (consistency index 0.987, retention index 0.981). Numbers below branches represent bootstrap values, taken from 1,000 replicates. FGSC4, NRRL 226, H640, NRRL 242, NRRL 244, NRRL 249 and ATCC 16813 are isolate numbers. GenBank accession numbers: *E. nidulans*, X02390; *A. versicolor*, AF058967; *A. sydowii* H640, AF058968; Saba, AF058969; San Salvador, AF058970; Florida Keys, AF058971; NRRL 242, AF058972; NRRL 244, AF058973; NRRL 249, AF058974; *E. violacea*, AF058975.

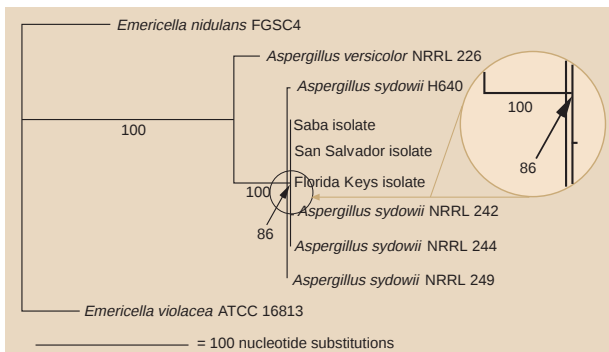


Table 1 Haplotypes classified by UEP groups

Haplotype	AC			SC		AL		SL		AI		SI	
	N	N	Freq.	N	Freq.	N	Freq.	N	Freq.	N	Freq.	N	Freq.
A NGACCT	207	47	0.959	50	0.877	16	0.302	19	0.679	42	0.618	32	0.628
B NGACTT	66	2	0.041	6	0.105	30	0.566	5	0.178	12	0.176	12	0.235
C PAACCT	33			1	0.018	7	0.132	4	0.143	14	0.206	7	0.137
Total	306	49	1.000	57	1.000	53	1.000	28	1.000	68	1.000	51	1.000
<i>h</i>		0.079		0.219		0.571		0.487		0.545		0.532	

For each sample, six microsatellites were typed (*DYS19*, *DYS388*, *DYS390*, *DYS391*, *DYS392* and *DYS393* (ref. 5)) and repeat counts determined. Y-chromosome haplotypes are listed in UEP groups (A–C) defined by six binary polymorphisms (YAP⁶, *SRY+4064* (ref. 7), *sY81* (ref. 8), *SRY+465* (T. Shinka and Y. Nakahori, personal communication), 92R7 (ref. 9) and Tāt¹⁰). Most Cohen chromosomes are included in a single UEP group (A), whereas both Levites and Israelites have a significant number in at least two groups. Within UEP group A, in both the Sephardic and the Ashkenazic Cohanim, a single haplotype, the Cohen modal haplotype, is dominant. The microsatellite scores for this haplotype, in the order of the microsatellite polymorphisms listed, are 14, 16, 23, 10, 11 and 12, according to the nomenclature of ref. 5. UEP groups are named by using references for the binary polymorphisms, in the order they are listed in parentheses above, as follows: P, YAP insert present; N, no YAP insert; A, adenine; C, cytosine; G, guanine; T, thymine. Priestly and lay classifications are: AC, Ashkenazic Cohen; SC, Sephardic Cohen; AL, Ashkenazic Levite; SL, Sephardic Levite; AI, Ashkenazic Israelite; SI, Sephardic Israelite. Ashkenazic: Jewish communities of northern Europe; Sephardic: Jewish communities of North Africa and the Middle East. *h* is a measure of genetic diversity¹¹. DNA was obtained by taking mouth swabs from 306 males in Israel, Canada and the UK. Before sample collection, prospective donors were questioned to ascertain the geographical origin of their fathers and grandfathers and their priestly/lay status; otherwise, donor selection was random. Some typings for the YAP polymorphism and for *DYS19* were first reported elsewhere¹. Polymerase chain reaction (PCR) primers for amplification of the microsatellites⁸ and biallelic markers^{8–10} were as described, except for *SRY+4064*, *sY81*, *DYS391* and *DYS392*, which were redesigned, and 92R7, which was converted to PCR format by M. Hurler and F. Santos in the laboratory of C. Tyler Smith.

from Israel, Canada and the United Kingdom. We found 112 compound haplotypes (see Supplementary information). Despite extensive diversity among Israelites, a single haplotype (the Cohen modal haplotype) is strikingly frequent in both Ashkenazic and Sephardic Cohanim (Table 1 and Supplementary information).

Because of microsatellite instability, it is useful to define a modal cluster of related chromosomes as the modal haplotype and all of its one-mutation neighbours at the microsatellite loci, which all share the same UEP markers. In the Ashkenazic and Sephardic Cohanim, the modal haplotype (cluster) frequencies are 0.449 (0.694) and 0.561 (0.614), respectively. For comparison, among the Ashkenazic and Sephardic Israelites, the frequencies are 0.132 (0.147) and 0.098 (0.138), respectively.

The Levites, unlike the Cohanim, have a significant number of Y chromosomes in three different UEP-defined groups (Table 1), indicating that Levite Y chromosomes have heterogeneous origins. Contemporary Levites, therefore, are not direct patrilineal descendants of a paternally related tribal group. Identification of the frequency of UEP group B chromosomes, and of the Ashkenazic Levite modal haplotype in particular (Supplementary information), in other populations may help in discovering the origins of Levite heterogeneity.

For highly polymorphic, single-locus systems, the identification of haplotypes with restricted distributions may provide 'signatures' of ancient connections that have been partially obscured by subsequent mixing with other populations. Gene flow from the Cohanim could account for the presence of the Cohen modal haplotype in both Ashkenazic and Sephardic Israelites, or it could be a signature of the ancient Hebrew population. The Cohen modal haplotype may therefore be useful for testing hypotheses regarding the relationship between specific contemporary communities and the ancient Hebrew population.

Given the relative isolation of Ash-

kenazic and Sephardic communities over the past 500 years, the presence of the same modal haplotype in the Cohanim of both communities strongly suggests a common origin. It is interesting, therefore, to estimate the time at which Cohen chromosomes were derived from a common ancestral chromosome (coalescence time). We assume that the modal haplotype is ancestral because of its high frequency, and we exclude from the analysis the few Cohen chromosomes that appear to be unrelated to it because of their membership of different UEP groups (Table 1). The distribution of allele sizes within UEP groups at the trinucleotide microsatellite *DYS388* indicates a departure from the stepwise mutation model. Because this model underlies the method we will use to estimate the coalescence time of Cohen chromosomes, we dropped *DYS388* from the analysis.

Under stepwise mutations, the average squared difference (ASD) in allele size among all current chromosomes and the ancestral haplotype, averaged over loci^{2,3}, has an expectation of μt , where μ is the mutation rate and t the coalescence time. Taking the Ashkenazic and Sephardic communities as a whole, the value for ASD is 0.2226. Assuming a mutation rate of 0.0021 (ref. 4), this gives an estimate of 106 generations, which for a generation time of 25 (30) years gives an estimate of 2,650 (3,180) years before present, dating the coalescence of the Cohanim chromosomes to between the Exodus and the destruction of the first Temple in 586 BC. Estimates based on the Ashkenazic and Sephardic samples taken separately are 2,619 (3,142) and 2,684 (3,221) years before present, respectively.

To obtain confidence intervals on the distance between the ancestral and sampled chromosomes (ignoring uncertainty in the mutation rate), we note that most non-ancestral haplotypes are singletons, indicating that the genealogy connecting Cohen chromosomes is more like the 'star genealogy' characteristic of rapid growth than the correlated genealogy characteristic of con-

stant size populations. To obtain confidence intervals in this case, we assume that M mutations occur during the 106 generations, with M being a Poisson random variable with parameter 106μ . The number of mutations increasing allele size (d) is drawn from a binomial distribution with parameters 0.5 and M (0.5 reflects size symmetry of mutations), leading to the distance $D = (2d - M)^2$. In a star genealogy, we have 485 (the number of loci multiplied by the sample size) observations of D .

Confidence intervals are obtained by repeating this process 1,000 times and taking the associated 2.5 and 97.5 percentiles, leading to a 95% confidence interval of 84–130 generations for the combined Ashkenazic and Sephardic samples or for a generation time of 25 years, 2,100–3,250 years before present. Under these assumptions, the 95% confidence interval places the origin of priestly Y chromosomes sometime during or shortly before the Temple period in Jewish history. Uncertainty in the mutation rate significantly broadens these intervals (conservatively taking 95% confidence intervals on both the distance and the mutation rate leads to an interval of 34–455 generations) as would a different assumption about the shape of the Cohen Y-chromosome genealogy.

Mark G. Thomas

The Centre for Genetic Anthropology,
Departments of Biology and Anthropology,
University College London, London WC1E 6BT, UK
Karl Skorecki*†, **Haim Ben-Ami**†

* Bruce Rappaport Faculty of Medicine and
Research Institute, Technion, Haifa 31096, Israel
† Rambam Medical Centre, Haifa 31096, Israel

Tudor Parfitt

School of Oriental and African Studies,
University of London, London WC1H 0XG, UK
Neil Bradman, **David B. Goldstein**
Department of Zoology, University of Oxford,
Oxford OX1 3PS, UK
e-mail: david.goldstein@zoo.ox.ac.uk

1. Skorecki, K. et al. *Nature* **385**, 32 (1997).

2. Goldstein, D. B., Linares, A. R., Cavalli-Sforza, L. L. & Feldman, M. W. *Genetics* **139**, 463–471 (1995).

3. Slatkin, M. A. *Genetics* **139**, 457–462 (1995).
4. Heyer, E., Puymirat, J., Deltjes, P., Bakker, E. & de Knijff, P. *Hum. Mol. Genet.* **6**, 799–803 (1997).
5. Kayser, M. et al. *Int. J. Legal Med.* **110**, 125–133, 141–149 (1997).
6. Hammer, M. F. *Mol. Biol. Evol.* **11**, 749–761 (1994).
7. Whitfield, L. S., Sulston, J. E. & Goodfellow, P. N. *Nature* **378**, 379–380 (1995).
8. Seielstad, M. T. et al. *Hum. Mol. Genet.* **3**, 2159–2161 (1994).
9. Mathias, N., Bayes, M. & Tyler-Smith, C. *Hum. Mol. Genet.* **3**, 115–123 (1994).
10. Zerjal, T. et al. *Am. J. Hum. Genet.* **60**, 1174–1183 (1997).
11. Nei, M. *Molecular Evolutionary Genetics* 177 (Columbia Univ. Press, New York, 1987).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Regeneration in metazoan larvae

The capacity for regeneration of lost body parts requiring organogenesis has never been described for metazoan larvae. Here we show for the first time, to our knowledge, the ability of sea star larvae to rapidly and completely regenerate to form fully functional individuals following surgical bisection of the larval body. This capacity leads us to extend the developmental stage at which echinoderm larval cell fates are irreversibly determined. As planktonic larval mortality is thought to be extremely high¹, this may be an adaptation to enhance recruitment into adult populations.

Larvae of the sea stars *Pisaster ochraceus* (bipinnaria and brachiolaria) and *Luidia foliolata* (bipinnaria) that had been surgically bisected across the horizontal larval axis with a scalpel, equidistant between the anterior and posterior poles, regenerated into fully functional larvae over a period of 12–14 days. No mortality occurred in any of the experimental or control larval treatments ($n=10$ –12 larvae per treatment). The ontogeny of larval regeneration was similar in both species and is shown for *P. ochraceus* in Fig. 1.

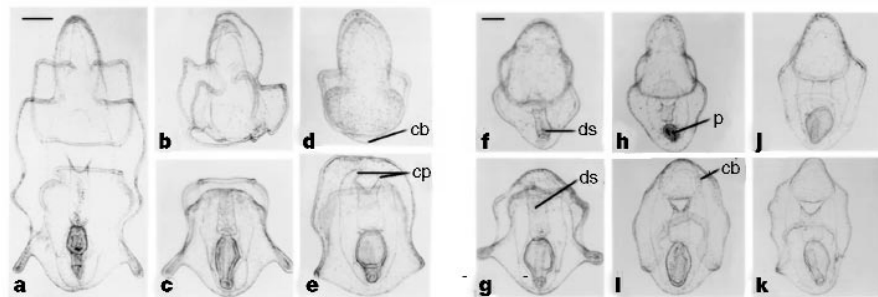


Figure 1 Bipinnaria larvae of *P. ochraceus* at various stages of regeneration after surgical bisection. The anterior larva is shown in the top row, and the posterior larva in the bottom row. **a**, Bipinnaria before surgical treatment. **b**, **c**, Day 1. Anterior (**b**) and posterior (**c**) larvae after surgical division. The gut of the posterior larva is intact enough for feeding. **d**, Day 3. **e**, Day 4. Elongation along the larval axis has taken place. **f**, **g**, Day 7. The anterior larva (**f**) has regenerated a functional digestive system. **h**, Day 9. **i**, Day 10. Complete regeneration of ciliary bands has occurred in both anterior and posterior larvae. **j**, **k**, Day 12. Both larvae have regenerated and are similar to control larvae. cb, ciliary band; cp, coelomic pouch; ds, digestive system; p, phytoplankton. **a**–**e**, Scale bar (in **a**), 200 μ m. **f**–**k**, Scale bar (in **f**), 200 μ m.

Immediately following surgical manipulation of intact larvae (Fig. 1a), both the anterior and the posterior portions of bipinnaria larvae (hereafter referred to as anterior and posterior larvae; Fig. 1b, c) exhibited swimming behaviours similar to those of control larvae. The posterior larvae consumed food, indicating that these larvae do not lose the ability to capture and ingest food particles, even though the upper portion of the oesophagus and mouth had been surgically removed.

By day 3, the anterior larvae had begun to regenerate the surgically separated postoral ciliary bands and to elongate along the larval axis (Fig. 1d). These larvae were able to capture and transport phytoplankton to the region of the former larval mouth. By day 4, the coelom of posterior larvae had elongated and showed signs of fusing above the larval mouth (Fig. 1e). After one week, both anterior and posterior larvae completed the regeneration of the digestive organs and coelom (Fig. 1f, g). At this point in the regenerative process, anterior larvae became able to feed. Capture and ingestion of phytoplankton were observed in culture; guts of anterior larvae contained phytoplankton by day 9 (Fig. 1h). The posterior larvae completely regenerated the pre-oral ciliary bands by day 10 (Fig. 1i).

By day 12, most anterior and posterior larvae had regenerated all of their body components (Fig. 1j, k), and surgically bisected bipinnaria larvae of both sea star species were essentially indistinguishable from control larvae by day 14.

Successive bisection of a bipinnaria larva of *P. ochraceus* resulted in complete regeneration indicative of sustained regenerative capacity. Brachiolaria larvae ($n=12$) showed similar chronological and morphological patterns of regeneration. In another study we bisected bipinnaria larvae of *L. foliolata* that had an adult rudiment ($n=12$). By day 3, rudiment-bearing posterior larvae had undergone complete metamorphosis into juveniles, whereas anterior and control

larvae required at least three to four times longer to metamorphose into juveniles.

Planktonic larvae of sea stars^{2–4} and brittle stars⁵ have been shown to reproduce asexually. Although this feature of echinoderm larvae reflects the indeterminate developmental nature of deuterostomal embryonic cells, such observations provide no direct measure of the ability of larvae to regenerate lost body components. Our observations of the ability of sea star larvae to regenerate show that we need to revise ideas about the stages of development of echinoderm larvae at which cell fates are considered to be irreversibly determined⁶.

High estimates of larval mortality are attributed to predation^{7,8}, abiotic factors, such as hydrodynamic shear stress⁹, or dispersal to inappropriate habitats¹⁰. Our results show that at least some larvae are capable of surviving and rapidly regenerating lost body parts following surgical bisection that simulates naturally induced damage. Therefore, population models that include rate functions of larval mortality in the plankton may be using mortality estimates that are unrealistically high.

It is also significant that damaged sea star larvae with an adult rudiment rapidly metamorphose to the benthic juvenile phase, thus reducing time spent exposed in the plankton and disproving hypotheses that resorption of the larval body is an energetic prerequisite to successful metamorphosis¹¹. Regenerative capacity may be particularly important in small, abundant, planktotrophic larvae that need to spend long periods of time developing in the plankton¹² and which exploit chemical defences less often than large, less abundant, yolk-laden lecithotrophic larvae^{13,14}.

Minako S. Vickery, James B. McClintock

Department of Biology,
University of Alabama at Birmingham,
Birmingham, Alabama 35294-1170, USA
e-mail: vickerym@uab.edu

1. Young, C. M. & Chia, F. S. in *Reproduction of Marine Invertebrates* (eds Giese, A. C., Pearse, J. S. & Pearse, V. B.) 385–463 (Boxwood, Pacific Grove, 1987).
2. Bosch, I., Rivkin, R. B. & Alexander, S. P. *Nature* **337**, 169–170 (1989).
3. Rao, P. S., Rao, K. H. & Shyam Sundari, K. A. *Curr. Sci.* **65**, 792–793 (1993).
4. Jaekle, W. B. *Biol. Bull.* **186**, 62–71 (1994).
5. Balser, E. J. *Am. Zool. Abstr.* **36**, 70 (1996).
6. Cameron, R. A., Hough-Evans, B. R., Britten, R. J. & Davidson, E. H. *Genes Dev.* **1**, 75–84 (1987).
7. Morgan, S. G. *J. Exp. Mar. Biol. Ecol.* **163**, 91–110 (1992).
8. Shanks, A. L. in *Ecology of Marine Invertebrate Larvae* (ed. McEdward, L. R.) 323–368 (CRC, Boca Raton, 1995).
9. Levitan, D. R. in *Ecology of Marine Invertebrate Larvae* (ed. McEdward, L. R.) 123–156 (CRC, Boca Raton, 1995).
10. Rummell, S. F. *Ophelia* **32**, 163–198 (1990).
11. Chia, F. S. & Burke, R. B. in *Settlement and Metamorphosis of Marine Invertebrate Larvae* (eds Chia, F. S. & Rice, M. E.) 212–234 (Elsevier, New York, 1978).
12. Emlet, R. B., McEdward, L. R. & Strathmann, R. R. in *Echinoderm Studies II* (eds Jangoux, M. & Lawrence, J. M.) 55–136 (Balkema, Rotterdam, 1987).
13. Lindquist, N. & Hay, M. E. *Ecol. Monogr.* **66**, 431–450 (1996).
14. McClintock, J. B. & Baker, B. J. *Mar. Ecol. Prog. Ser.* **154**, 121–131 (1997).

Postmodernism disrobed

Intellectual Impostures

by Alan Sokal and Jean Bricmont

Profile: 1998. Pp. 274. £9.99

To be published in the USA by Picador as Fashionable Nonsense in November 1998

Richard Dawkins

Suppose you are an intellectual impostor with nothing to say, but with strong ambitions to succeed in academic life, collect a coterie of reverent disciples and have students around the world anoint your pages with respectful yellow highlighter. What kind of literary style would you cultivate? Not a lucid one, surely, for clarity would expose your lack of content. The chances are that you would produce something like the following:

We can clearly see that there is no bi-univocal correspondence between linear signifying links or archi-writing, depending on the author, and this multireferential, multi-dimensional machinic catalysis. The symmetry of scale, the transversality, the pathic non-discursive character of their expansion: all these dimensions remove us from the logic of the excluded middle and reinforce us in our dismissal of the ontological binarism we criticised previously.

This is a quotation from the psychoanalyst Félix Guattari, one of many fashionable French 'intellectuals' outed by Alan Sokal and Jean Bricmont in their splendid book *Intellectual Impostures*, previously published in French and now released in a completely rewritten and revised English edition. Guattari goes on indefinitely in this vein and offers, in the opinion of Sokal and Bricmont, "the most brilliant mélange of scientific, pseudo-scientific and philosophical jargon that we have ever encountered". Guattari's close collaborator, the late Gilles Deleuze, had a similar talent for writing:

In the first place, singularities-events correspond to heterogeneous series which are organized into a system which is neither stable nor unstable, but rather 'metastable', endowed with a potential energy wherein the differences between series are distributed... In the second place, singularities possess a process of auto-unification, always mobile and displaced to the extent that a paradoxical element traverses the series and makes them resonate, enveloping the corresponding singular points in a single aleatory point and all the emissions, all dice throws, in a single cast.

This calls to mind Peter Medawar's earlier characterization of a certain type of French intellectual style (note, in passing, the contrast offered by Medawar's own elegant and clear prose):

Style has become an object of first importance, and what a style it is! For me it has a



How shall we know whether the modish French 'philosophy'... is genuinely profound or the vacuous rhetoric of mountebanks and charlatans?

prancing, high-stepping quality, full of self-importance; elevated indeed, but in the balletic manner, and stopping from time to time in studied attitudes, as if awaiting an outburst of applause. It has had a deplorable influence on the quality of modern thought...

Returning to attack the same targets from another angle, Medawar says:

I could quote evidence of the beginnings of a whispering campaign against the virtues of clarity. A writer on structuralism in the Times Literary Supplement has suggested that thoughts which are confused and tortuous by reason of their profundity are most appropriately expressed in prose that is deliberately unclear. What a preposterously silly idea! I am reminded of an air-raid warden in wartime Oxford who, when bright moonlight seemed to be defeating the spirit of the blackout, exhorted us to wear dark glasses. He, however, was being funny on purpose.

This is from Medawar's 1968 lecture on "Science and Literature", reprinted in *Pluto's Republic* (Oxford University Press, 1982). Since Medawar's time, the whispering campaign has raised its voice.

Deleuze and Guattari have written and collaborated on books described by the celebrated Michel Foucault as "among the greatest of the great... Some day, perhaps, the century will be Deleuzian." Sokal and Bricmont, however, think otherwise: "These texts contain a handful of intelligible sentences — sometimes banal, sometimes erroneous — and we have commented on some of them in

the footnotes. For the rest, we leave it to the reader to judge."

But it's tough on the reader. No doubt there exist thoughts so profound that most of us will not understand the language in which they are expressed. And no doubt there is also language designed to be unintelligible in order to conceal an absence of honest thought. But how are we to tell the difference? What if it really takes an expert eye to detect whether the emperor has clothes? In particular, how shall we know whether the modish French 'philosophy', whose disciples and exponents have all but taken over large sections of American academic life, is genuinely profound or the vacuous rhetoric of mountebanks and charlatans?

Sokal and Bricmont are professors of physics at, respectively, New York University and the University of Louvain in Belgium. They have limited their critique to those books that have ventured to invoke concepts from physics and mathematics. Here they know what they are talking about, and their verdict is unequivocal. On Jacques Lacan, for example, whose name is revered by many in humanities departments throughout US and British universities, no doubt partly because he simulates a profound understanding of mathematics:

...although Lacan uses quite a few key words from the mathematical theory of compactness, he mixes them up arbitrarily and without the slightest regard for their meaning. His 'definition' of compactness is not just false: it is gibberish.

They go on to quote the following remarkable piece of reasoning by Lacan:

Thus, by calculating that signification according to the algebraic method used here, namely:

$$\frac{S(\text{signifier})}{s(\text{signified})} = s(\text{the statement}),$$

with $S=(1)$, produces: $s=\sqrt{-1}$.

You don't have to be a mathematician to see that this is ridiculous. It recalls the Aldous Huxley character who proved the existence of God by dividing zero into a number, thereby deriving the infinite. In a further piece of reasoning that is entirely typical of the genre, Lacan goes on to conclude that the erectile organ

...is equivalent to the $\sqrt{-1}$ of the signification produced above, of the jouissance that it restores by the coefficient of its statement to the function of lack of signifier (-1) .

We do not need the mathematical expertise of Sokal and Bricmont to assure us that the author of this stuff is a fake. Perhaps he is genuine when he speaks of non-scientific subjects? But a philosopher who is caught equating the erectile organ to the square root of minus one has, for my money, blown his

credentials when it comes to things that I don't know anything about.

The feminist 'philosopher' Luce Irigaray is another who gets whole-chapter treatment from Sokal and Bricmont. In a passage reminiscent of a notorious feminist description of Newton's *Principia* (a "rape manual"), Irigaray argues that $E=mc^2$ is a "sexed equation". Why? Because "it privileges the speed of light over other speeds that are vitally necessary to us" (my emphasis of what I am rapidly coming to learn is an 'in' word). Just as typical of this school of thought is Irigaray's thesis on fluid mechanics. Fluids, you see, have been unfairly neglected. "Masculine physics" privileges rigid, solid things. Her American expositor Katherine Hayles made the mistake of re-expressing Irigaray's thoughts in (comparatively) clear language. For once, we get a reasonably unobstructed look at the emperor and, yes, he has no clothes:

The privileging of solid over fluid mechanics, and indeed the inability of science to deal with turbulent flow at all, she attributes to the association of fluidity with femininity. Whereas men have sex organs that protrude and become rigid, women have openings that leak menstrual blood and vaginal fluids... From this perspective it is no wonder that science has not been able to arrive at a successful model for turbulence. The problem of turbulent flow cannot be solved because the conceptions of fluids (and of women) have been formulated so as necessarily to leave unarticulated remainders.

You do not have to be a physicist to smell out the daffy absurdity of this kind of argument (the tone of it has become all too familiar), but it helps to have Sokal and Bricmont on hand to tell us the real reason why turbulent flow is a hard problem: the Navier-Stokes equations are difficult to solve.

In similar manner, Sokal and Bricmont expose Bruno Latour's confusion of relativity with relativism, Jean-François Lyotard's 'post-modern science', and the widespread and predictable misuses of Gödel's Theorem, quantum theory and chaos theory. The renowned Jean Baudrillard is only one of many to find chaos theory a useful tool for bamboozling readers. Once again, Sokal and Bricmont help us by analysing the tricks being played. The following sentence, "though constructed from scientific terminology, is meaningless from a scientific point of view":

Perhaps history itself has to be regarded as a chaotic formation, in which acceleration puts an end to linearity and the turbulence created by acceleration deflects history definitively from its end, just as such turbulence distances effects from their causes.

I won't quote any more, for, as Sokal and Bricmont say, Baudrillard's text "continues in a gradual crescendo of nonsense". They again call attention to "the high density of scientific and pseudo-scientific terminology — inserted in sentences that are, as far as we



MARK DOIRON

This was a *physicist* saying all the right-on things they wanted to hear, attacking the 'post-Enlightenment hegemony' and such uncool notions as the existence of the real world.

can make out, devoid of meaning". Their summing up of Baudrillard could stand for any of the authors criticized here and lionized throughout America:

In summary, one finds in Baudrillard's works a profusion of scientific terms, used with total disregard for their meaning and, above all, in a context where they are manifestly irrelevant. Whether or not one interprets them as metaphors, it is hard to see what role they could play, except to give an appearance of profundity to trite observations about sociology or history. Moreover, the scientific terminology is mixed up with a non-scientific vocabulary that is employed with equal sloppiness. When all is said and done, one wonders what would be left of Baudrillard's thought if the verbal veneer covering it were stripped away.

But don't the postmodernists claim only to be 'playing games'? Isn't the whole point of their philosophy that anything goes, there is no absolute truth, anything written has the

same status as anything else, and no point of view is privileged? Given their own standards of relative truth, isn't it rather unfair to take them to task for fooling around with word games, and playing little jokes on readers? Perhaps, but one is then left wondering why their writings are so stupefyingly boring. Shouldn't games at least be entertaining, not po-faced, solemn and pretentious? More tellingly, if they are only joking, why do they react with such shrieks of dismay when somebody plays a joke at their expense? The genesis of *Intellectual Impostures* was a brilliant hoax perpetrated by Sokal, and the stunning success of his coup was not greeted with the chuckles of delight that one might have hoped for after such a feat of deconstructive game playing. Apparently, when you've become the establishment, it ceases to be funny when someone punctures the established bag of wind.

As is now rather well known, in 1996 Sokal submitted to the US journal *Social Text* a paper called "Transgressing the boundaries: towards a transformative hermeneutics of quantum gravity". From start to finish the paper was nonsense. It was a carefully crafted parody of postmodern metatwaddle. Sokal was inspired to do this by Paul Gross and Normal Levitt's *Higher Superstition: The Academic Left and its Quarrels with Science* (Johns Hopkins University Press, 1994), an important book that deserves to become as well known in Britain as it is in the United States. Hardly able to believe what he read in

this book, Sokal followed up the references to postmodern literature, and found that Gross and Levitt did not exaggerate. He resolved to do something about it. In the words of the journalist Gary Kamiya:

Anyone who has spent much time wading through the pious, obscurantist, jargon-filled cant that now passes for 'advanced' thought in the humanities knew it was bound to happen sooner or later: some clever academic, armed with the not-so-secret passwords ('hermeneutics,' 'transgressive,' 'Lacanian,' 'hegemony,' to name but a few) would write a completely bogus paper, submit it to an au courant journal, and have it accepted... Sokal's piece uses all the right terms. It cites all the best people. It whacks sinners (white men, the 'real world'), applauds the virtuous (women, general meta-physical lunacy)... And it is complete, unadulterated bullshit — a fact that somehow escaped the attention of the high-powered editors of Social Text, who must now be experiencing that queasy sensation that afflicted the Trojans the morning after they pulled that nice big gift horse into their city.

Sokal's paper must have seemed a gift to the editors because this was a *physicist* saying all the right-on things they wanted to hear, attacking the 'post-Enlightenment hegemony' and such uncool notions as the existence of the real world. They didn't know that Sokal had also crammed his paper with egregious scientific howlers, of a kind that any referee with an undergraduate degree in physics would instantly have detected. It was sent to no such referee. The editors, Andrew Ross and others, were satisfied that its ideology conformed to their own, and were perhaps flattered by references to their own works. This ignominious piece of editing rightly earned them the 1996 Ig Nobel prize for literature.

Notwithstanding the egg all over their faces, and despite their feminist pretensions, these editors are dominant males in the academic establishment. Ross has the boorish, tenured confidence to say things like, "I am glad to be rid of English departments. I hate literature, for one thing, and English departments tend to be full of people who love literature"; and the yahooish complacency to begin a book on 'science studies' with these words: "This book is dedicated to all of the science teachers I never had. It could only have been written without them."

He and his fellow 'cultural studies' and 'science studies' barons are not harmless eccentrics at third-rate state colleges. Many of them have tenured professorships at some of the best universities in the United States. Men of this kind sit on appointment committees, wielding power over young academics who might secretly aspire to an honest academic career in literary studies or, say, anthropology. I know — because many of them have told me — that there are sincere scholars out there who would speak out if

they dared, but who are intimidated into silence. To them, Sokal will appear as a hero, and nobody with a sense of humour or a sense of justice will disagree. It helps, by the way, although it is strictly irrelevant, that his own left-wing credentials are impeccable.

In a detailed post-mortem of his famous hoax, submitted to *Social Text* but predictably rejected by them and published elsewhere, Sokal notes that, in addition to numerous half-truths, falsehoods and *non sequiturs*, his original article contained some "syntactically correct sentences that have no meaning whatsoever". He regrets that there were not more of these: "I tried hard to produce them, but I found that, save for rare bursts of inspiration, I just didn't have the knack." If he were writing his parody today, he would surely be helped by a virtuoso piece of computer programming by Andrew Bulhak of Melbourne, Australia: the Postmodernism Generator. Every time you visit it, at <http://www.cs.monash.edu.au/cgi-bin/postmodern>, it will spontaneously generate for you, using faultless grammatical principles, a spanking new postmodern discourse, never before seen.

I have just been there, and it produced for me a 6,000-word article called "Capitalist theory and the subtextual paradigm of context" by "David I. L. Werther and Rudolf du Garbandier of the Department of English, Cambridge University" (poetic justice there, for it was Cambridge that saw fit to give Jacques Derrida an honorary degree). Here is a typical passage from this impressively erudite work:

If one examines capitalist theory, one is faced with a choice: either reject neotextual materialism or conclude that society has objective value. If dialectic desituationism holds, we have to choose between Habermasian discourse and the subtextual paradigm of context. It could be said that the subject is contextualised into a textual nationalism that includes truth as a reality. In a sense, the premise of the subtextual paradigm of context states that reality comes from the collective unconscious.

Visit the Postmodernism Generator. It is a literally infinite source of randomly generated, syntactically correct nonsense, distinguishable from the real thing only in being more fun to read. You could generate thousands of papers per day, each one unique and ready for publication, complete with numbered endnotes. Manuscripts should be submitted to the 'Editorial Collective' of *Social Text*, double-spaced and in triplicate.

As for the harder task of reclaiming US literary departments for genuine scholars, Sokal and Bricmont have joined Gross and Levitt in giving a friendly and sympathetic lead from the world of science. We must hope that it will be followed. □

Richard Dawkins is at the Oxford University Museum of Natural History, Parks Road, Oxford OX1 3PW, UK.

From protective rind to cognitive cortex

Brain, Vision, Memory: Tales in the History of Neuroscience

by Charles G. Gross

MIT Press: 1998. Pp. 255. \$30, £19.95

John C. Marshall

John Maynard Keynes wrote: "I do not know what makes a man more conservative — to know nothing but the present, or nothing but the past." Happily, Charles Gross is in no danger of being impaled on either horn of Keynes's aphorism. A busy experimental neuroscientist, Gross first came to fame as the discoverer of cells in the macaque's inferotemporal cortex that respond selectively to the image of a monkey's hand — not quite 'grandmother cells' (which would respond only to the face of one's grandmother), but close enough. More recently, he has found neurons in the central pre-motor cortex of the monkey that respond to both tactile and visual stimuli in the space adjacent to the animal's arm and hand.

I mention this background to indicate that Gross is not a historian *per se*. His longstanding (but relatively newly indulged) interest in the development of his discipline over the past three or four millennia is that of an active practitioner, not a disinterested embalmer. From the Egyptian neurosurgeons of the Old Kingdom to David Ferrier, by way of Andreas Vesalius of Padua, *Brain, Vision, Memory* treats the distinguished figures of the past as if they were our contemporaries, struggling as we do to make sense of how all perception, thought and practice depends on the functioning of an organ that looks more like a huge walnut than the seat of the soul.

Gross's tales of the history of neuroscience can be warmly recommended to all students of the brain, but especially to those who believe that history began when they were undergraduates. Informative and amusing in equal part, Gross is as fair to those who were wildly wrong as to those who were (relatively) right. This is just as well: human brains each contain 10^{11} neurons and 10^{14} synapses (at least), and it is accordingly unlikely that our rather puny minds will ever fully understand the massive structures that give rise to their associated psyches.

But this is not for want of trying. In the first and longest essay in the book, Gross tells the story of visual cortex "from Imhotep to Hubel and Wiesel". The Edwin Smith surgical papyrus (from around 1700 BC) provides the first documentation that brain and behaviour are related to each other, yet the prevalence of injuries antipodal to the causative blow blocked the insight that each side of the brain controls the opposite side of the body. Once the brain was conjectured to



Believing is seeing: here, even Leonardo drew what he thought he knew, rather than what he saw.

be the crucial organ of cognition, the search was on for assigning different functions to distinct regions. Are the ventricles the substrate of computation, and the cerebral cortex mere protective covering? If the cortex is the seat of 'higher' functions, are those offices equipotentially distributed throughout the cortex? If there is a brain 'centre' for vision, is it localized in parietal or in occipital cortex?

Gross expertly leads the reader through the highways and byways of how and why better answers to such questions become available. In a story that is never less than fascinating, a few gems stand out. It is well known that Aristotle argued that the brain plays a subordinate role in the heart-brain system responsible for sensation, intelligence and movement. Gross's account of how Aristotle could find this view both rationally plausible and empirically valid is superb. Likewise, Gross performs a mitzvah in remembering the early brilliance, mature achievement and tragic fate of the Polish sensory neurophysiologist Adolf Beck.

Another essay, primarily concerned with the visual system, discusses Leonardo da Vinci's neuroanatomical illustrations. The chapter is a salutary reminder that not even the greatest of draughtsmen could entirely avoid drawing what he (thought he) knew, rather than what he actually saw. An intriguing digression on the irony of how "Leonardo, who presumably easily read his own left-right reversed writing, found it inconceivable that the brain could interpret an inverted image" is all too short.

The fame of Leonardo provides a striking contrast to the obscurity (within neuroscience, at least) of another all-round genius, Emanuel Swedenborg, to whom Gross devotes a very welcome chapter. Swedenborg "read widely about the brain in the biological and medical literature of the day". His

extensive writings on the brain and sense organs turn out to be observationally more accurate and theoretically more profound than those of any other eighteenth-century scholar. Gross attributes the neglect of Swedenborg to the fact that he failed to publish (in relevant places). This is no doubt correct, and should serve as a dire warning to us all. But one also suspects that Swedenborg's invention of a new religion — the New Jerusalem Church — counted against him when research income per head of staff was calculated: the Swedish government paid him to be Inspector of Mines, not to worry about the biological bases of the soul.

The final chapter (on how vision purportedly took over more and more of the brain) takes us from the end of the nineteenth century (Paul Fleschig's work on visual association cortex) to the 'what' and 'where' pathways of current neuroscience. Although Gross can still find mistakes aplenty in this period, even he finds it hard to suppress the urge to rejoice that now, at last, we've got it right.

The reader should keep the dust-jacket, on which Torsten Wiesel writes of his hope that Gross's magnificent volume will "encourage today's neuroscientists to search carefully for their own misconceptions and follies". Gross concludes with a warning against the reification of neuropsychological categories. But the stories that he tells suggest that the reification of neuroanatomical structures is a greater danger. □

John C. Marshall is at the Neuropsychology Unit, University Department of Clinical Neurology, Radcliffe Infirmary, Woodstock Road, Oxford OX2 6HE, UK.

A piece of the true Feynman

The Meaning of it All: Thoughts of a Citizen Scientist

by Richard P. Feynman

Addison Wesley: 1998. Pp.122. £12.99, \$22

Stephen Battersby

A small volume has appeared that transcribes a lecture series given by Richard Feynman. This is not an unfamiliar happening; new bits of Feynmania appear quite frequently, and I suppose we should be as used to the cult of Feynman as we are to other celebrity cults. Then again, because Feynman is set up as a thinker and guru, even as a prophet, perhaps his status should be more carefully examined.

Here he sets out to tackle science's relations with politics, religion and everyday society. If Feynman was a prophet, I suppose this was his sermon on divine uncertainty — the uncertainty that allows the scientific process to work. Knowledge can progress, says Feynman, only if people have open

minds and test their ideas. So far so good.

Not surprisingly, dogma comes in for a bashing. Feynman is relatively gentle with religion, only pointing out that it is hard to reconcile a culture of doubt (science) with one of faith. He is much fiercer with political dogma, especially with the communism of the former Soviet Union, so much less friendly to new ideas than the uncertainty-enshrining democracy of the United States. Although the words he uses are measured, he admits to becoming over-emotional at this point. Well, this was 1963, only months after the Cuban missile crisis, so perhaps it is understandable.

But then we hit a snag: in print, and unedited, Feynman doesn't always make sense. In one crucial passage he tries to argue that "ethical values lie outside the scientific realm". This must be a comforting opinion for someone who worked on the bomb, but he is on controversial ground, and here most of all he reveals himself to be surprisingly inarticulate. Here is one whole argument for the separation of science and ethics: "First, in the past there were conflicts. The metaphysical positions have changed, and there have [sic] been practically no effect on the ethical views. So there must be a hint that there is an independence." Eh? Well, I catch the drift, but one would hope for more than that. Ordinary people have this sort of vague insight all the time; what one wants from a guru is a little more intellectual clarity — or at least more memorable phrasing.

It does get better. In the last lecture, he has a go at the sort of everyday stupidity that comes of not thinking scientifically, especially from having a poor sense of probability. People justify their beliefs in flying saucers, astrology and faith healing, for example, because these things are "possible". These may be soft targets but they always need attacking, and it is fun to watch them being hit so hard. Even Feynman's arrogance can be endearing here, and one can imagine his tone of exasperation when he said "the number of things that are possible is not fully appreciated by the average individual".

The Feynman of these lectures is certainly sensible and amusing, but not inspirational. Don't visit him for sacred wisdom. □
Stephen Battersby is an assistant editor of *Nature*.



Feynman: surprisingly inarticulate.

Three-dimensional structure of the Stat3 β homodimer bound to DNA

Stefan Becker*, Bernd Groner† & Christoph W. Müller*

* European Molecular Biology Laboratory, Grenoble Outstation, c/o ILL BP 156, 38042 Grenoble Cedex 9, France

† Institute for Experimental Cancer Research, Tumor Biology Center, Breisacher Strasse 117, D-79106 Freiburg i. Br., Germany

STAT proteins are a family of eukaryotic transcription factors that mediate the response to a large number of cytokines and growth factors. Upon activation by cell-surface receptors or their associated kinases, STAT proteins dimerize, translocate to the nucleus and bind to specific promoter sequences on their target genes. Here we report the first crystal structure of a STAT protein bound to its DNA recognition site at 2.25 Å resolution. The structure provides insight into the various steps by which STAT proteins deliver a response signal directly from the cell membrane to their target genes in the nucleus.

STAT (for signal transducers and activators of transcription) proteins are activated in response to extracellular signals such as cytokines, growth factors and hormonal factors. Binding of these factors to cell-surface receptors leads to receptor autophosphorylation at a tyrosine residue. STAT proteins, which contain an SH2 domain that recognizes the receptor phosphotyrosine, are recruited from the cytosol and associate with the activated receptor. They then become phosphorylated on a carboxy-terminal tyrosine, either directly by the receptor or by a receptor-associated JAK kinase. Tyrosine phosphorylation of STATs causes them to dimerize and translocate to the nucleus, where they bind to specific promoter sequences in target genes (reviewed in refs 1–3).

Seven different STAT genes have been identified so far in mammals. The encoded proteins vary in length between 750 and 850 amino acid residues and share 20 to 50% pairwise sequence

identity. In general, STAT proteins bind as dimers to DNA target sites with a 9-base-pair (bp) consensus sequence, TTCCGGGAA, and binding constants in the nanomolar range. Additional roles of monomeric Stat1 in the gene regulation of apoptosis⁴ and of monomeric Stat3 as an adaptor for the recruitment of phosphatidylinositol-3-OH kinase to its receptor⁵ have been described.

Sequence comparisons, biochemical assays and mutagenesis have identified distinct functional domains within the STAT molecules (Fig. 1a). The 130 N-terminal amino-acid residues mediate cooperativity in binding to multiple DNA sites^{6,7}. Residues 400–500 confer DNA-binding specificity but are not sufficient for DNA binding⁸. Residues 600–700 share homology with Src-homology-2 (SH2) domains and mediate dimerization as a result of phosphotyrosine recognition⁹. The phosphorylated tyrosine is located around residue 700 (Tyr705 in Stat3). The carboxy terminus is

Table 1 Structure determination of the Stat3 β –DNA complex

Data collection					
Data set	Resolution range (Å)	R_{sym} (%)*	ESRF beam line	Reflection observed/unique	Completeness (%)
Native†	20–2.25	5.3	ID2	273,455/56,005	99.4
MMA‡	20–2.45	7.0	ID2	206,659/42,055	97.0
PCMB§	15–2.9	8.9	ID2	107,982/25,777	98.8
Se-methionine	20–2.6	8.5	BM14	110,453/35,122	96.7
EMTS	20–3.2	7.6	ID14-3	33,123/15,318	81.6
K ₂ PtCl ₄	20–3.1	6.0	ID2	98,729/21,160	98.5
Phasing statistics					
Data set	Isomorphous difference (%)	Phasing power acentric/centric	R_{cullis} acentric/centric	Number of sites	Figure of merit acentric/centric
Native	—	—	—	—	0.555/0.599
MMA	25.1	2.29/1.96	0.64/0.69	8	
PCMB§	21.0	2.70/2.00	0.59/0.65	9	
Se-methionine	9.4	0.73/0.62	0.91/0.92	16	
EMTS	26.1	2.69/1.92	0.65/0.67	7	
K ₂ PtCl ₄	20.5	0.59/0.61	0.98/1.03	3	
Refinement					
Resolution range (Å)	20–2.25				
Total number of non-hydrogen atoms	4,985				
Number of water molecules	127				
R-factor (%)	25.4 (for 53,097 reflections)				
R_{free} (%)¶	30.6 (for 2,881 reflections)				
R.m.s. deviation					
Bond length (Å)	0.013				
Bond angles (deg)	2.1				

* $R_{\text{sym}} = \sum_{i,j} |I(i,h) - I(j,h)| / \sum_{i,j} I(i,h)$ where $I(h)$ is the mean of the i observations of reflection h .

† R_{sym} between 2.33 and 2.25 Å is 30.0%.

‡ Soaking conditions were 1 mM for 1 d for MMA (methylmercuryacetate) and PCMB§ (p-chloromercuribenzenesulphonate), 0.01 mM for 6 d for EMTS (ethylmercurithiosalicylate), and 0.05 mM for 2 d for K₂PtCl₄.

§ Isomorphous difference is $\Sigma |F_{\text{PH}} - F_{\text{P}}| / \Sigma F_{\text{PH}}$, F_{PH} and F_{P} are the derivative and native structure factor amplitudes, respectively.

|| Phasing power, mean value of heavy-atom structure factor amplitudes divided by the lack of closure. R_{cullis} , mean lack of closure divided by mean isomorphous difference; statistics calculated with program SHARP³⁷.

¶ R_{free} , calculated from a subset of 5% of the data.

important for transcriptional activation, which can be regulated by serine phosphorylation¹⁰. This region varies in its sequence and length among different homologues. The Stat3 β protein used for our structural analysis is a truncated isoform of Stat3 α , in which 55 C-terminal amino-acid residues are replaced by a stretch of 7 amino acid residues^{11,12}.

The crystal structure of the amino-terminal domain of Stat4 has been reported¹³. The C-terminal fragment lacking this domain still retains full binding activity to single DNA-target sites. On the basis of the results of Vinkemeier *et al.*⁷, we expressed the 65K C-terminal fragment of STAT3 β (residues 127 to 722) in bacteria; this fragment was then tyrosine-phosphorylated in bacteria. The phosphorylation occurred almost exclusively at Tyr 705 (data not shown) and caused the STAT protein to dimerize and bind to specific DNA oligonucleotides. We crystallized the activated protein in complex with a

17-base pair DNA duplex with overhanging ends and containing the high-affinity DNA target site M67 (ref. 14). Crystals diffracted to 2.2 Å resolution using synchrotron radiation. The structure was solved by multiple isomorphous replacement using mercury, platinum and selenomethionine derivatives, and was refined to an *R*-factor of 25.4% (*R*_{free} = 30.6%) using data between 20 and 2.25 Å resolution. The crystallographic statistics are summarized in Table 1.

Overall structure

The crystals of the Stat3 β homodimer:DNA complex (spacegroup *I*₄, *a* = *b* = 174.0 Å, *c* = 79.4 Å) contain half a complex per asymmetric unit. Our model comprises residues 136 to 716 of one monomer, half a DNA duplex, and 127 water molecules per asymmetric unit. A few residues at the N and C termini (residues 127–135, 717–722), the loop connecting helices α 1 and α 2 (residues 184–194) and residues 688 to 701, which link the SH2 domain with the phosphotyrosine-containing region (residues 702 to 716), are poorly ordered and are not included in our model. Residue 136 is the first residue with clear electron density. It lies close to the DNA and is preceded by the N-terminal domain.

The two monomers are related by a crystallographic dyad. Although the DNA duplex used for co-crystallization deviates from strict twofold symmetry at four base pairs (bp $\pm$ 8, bp $\pm$ 7, bp $\pm$ 2, bp 0) and the overlapping 5' end (Fig. 3a), its pseudodyad coincides with the crystallographic dyad. This is a consequence of the random bimodal orientation of the complex around the dyad, which was confirmed as described in the Methods.

As a result, our structure describes the crystallographic average of two halves of the complex bound to slightly different half-sites. However, this does not impair the interpretation of the structure: the experimental and the final $2F_o - F_c$ electron density of the DNA is of excellent quality for the sugar-phosphate backbone and the dimerically related bases, and is consistent with the superposition of two different base pairs at the non-palindromic positions. Only one of these positions (bp $\pm$ 2) is directly contacted by the protein, whereas one base pair (bp $\pm$ 0) forms a water-mediated contact (see below). The other two base pairs (bp $\pm$ 8, bp $\pm$ 7) lie outside the DNA consensus and are not contacted by the protein. All protein residues involved in interactions with the DNA show electron density of high quality and we do not observe any diffuse density indicative of residues at the interface that deviate from the crystallographic symmetry. This suggests that each monomer undergoes only minor changes to adapt to its half-sites and shows that the interpretation of the protein-DNA interface structure is not obscured by the random orientation of the complex around the dyad.

Although the electron density is of high quality for most of the model, it is considerably less well defined for the SH2 domain and the phosphopeptide region. This presumably reflects a higher degree of mobility in this region, which is consistent with the molecular packing in the crystal lattice. Nonetheless, structural similarity to other SH2 domains, reasonable stereochemistry for most residues, the ability to account for four selenomethionine and mercury markers, and the relatively favourable appearance of our omit maps makes us confident that our tracing of this region is essentially correct. Moreover, possible errors in this region (see Methods), do not affect our general conclusions regarding dimerization or phosphotyrosine binding.

The overall dimensions of the homodimer bound to DNA are 145 Å $\times$ 115 Å $\times$ 60 Å. The longest dimension is perpendicular to the DNA axis, whereas the shortest runs parallel to the DNA axis (Fig. 2). Each monomer is composed of four domains: an N-terminal four-helix bundle (residues 138–320), an eight-stranded β -barrel (residues 321–465), an α -helical 'connector' domain (residues 466–585) and an SH2 domain (residues 586–688). The phosphotyrosine occurs in a stretch of ordered residues at the C terminus (residues 702–716).

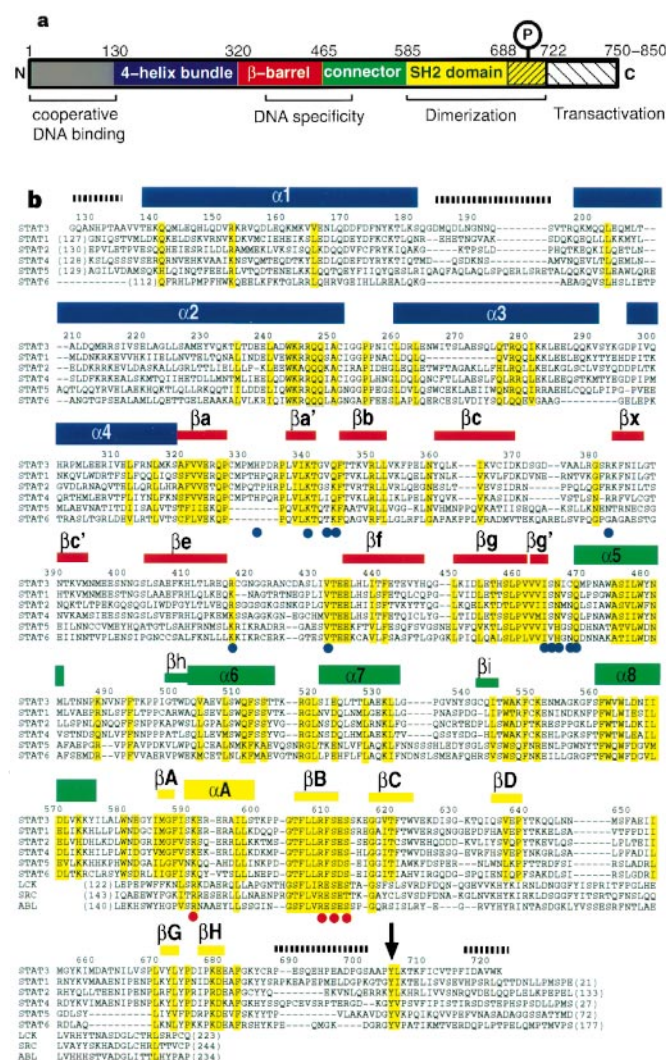


Figure 1 Structure of STAT proteins. **a**, The domain structure: the domain boundaries of Stat3 β are shown. The construct used for co-crystallization comprises residues 127–722 and lacks the N-terminal cooperativity domain as well as the C-terminal transactivating domain. **b**, Sequence alignment of mouse Stat3 β with human homologues Stat1, Stat2, Stat4, Stat5 and Stat6 and Lck, Src and Abl SH2 domains (program CLUSTALW, manually adjusted)⁴³. Conserved and conservatively substituted residues are highlighted in yellow. Secondary-structure elements follow the colour scheme of Fig. 2; strands in the β -barrel follow the convention used for the NF- κ B/NF-AT superfamily⁴⁴. Disordered regions are indicated by hatched lines. Arrow, phosphotyrosine residue. Residues involved in DNA and phosphotyrosine binding are marked by blue and red dots, respectively.

The STAT homodimer grips the DNA like a pair of pliers. The monomers are held together by the C-terminal SH2 domains, and the large 4-helix-bundle domains form the 'handles' of the pliers. The DNA is almost entirely enclosed by the protein dimer, in contact with loops from the β -barrel and the connector domains.

Structure of the protein

The N-terminal domain, a bundle of four antiparallel helices connected by short loops, is very elongated (~ 80 Å long). Helices $\alpha 1$ (44 residues) and $\alpha 2$ (54 residues) span the entire domain; helices $\alpha 3$ and $\alpha 4$ are considerably shorter (32 and 25 residues).

Helix $\alpha 2$, which has a kink at residue 234, packs against strand βa and helices $\alpha 5$ and $\alpha 6$, linking the 4-helix bundle with the β -barrel and the connector domain. The central hydrophobic core of the bundle contains Val, Leu and Ile residues. The underlying heptad spacing in this region was recognized by Fu *et al.*¹⁵, who correctly predicted a coiled-coil motif.

The 4-helix bundle is immediately followed by the eight-stranded β -barrel. Strands a, b and e form one sheet of the barrel; strands c, f and g form the other. Strands x/c' connect the two sheets by hydrogen bonding, first to strand e of sheet abe, then switching over to pair with strand c of sheet cfg. Strand extensions a' and g'

form a small two-stranded sheet closing the β -barrel at the end next to the DNA, where three loops of this domain participate in DNA binding (Figs 2, 3a, b).

The β -barrel domain is linked to the SH2 domain by a small helical domain, formed by two helix-loop-helix modules (helices $\alpha 5$ to $\alpha 8$), which we call the 'connector' domain. This domain shows structural similarity to calcium-binding domains, as found in troponin C^{16,17}, for example. In common with Ca²⁺-binding domains, the connector domain has an identical arrangement of helices, a pseudodyad between the $\alpha 5/\alpha 6$ and $\alpha 7/\alpha 8$ helix-loop-helix motifs (r.m.s._{34Ca-atoms} = 1.9 Å), and a small two-stranded β -pleated sheet, formed by strands h and i, pairing these two motifs. However, the loops of the connector domain are longer, and the polar residues that chelate the metal in calcium-binding domains are replaced by hydrophobic residues which pack against each other. The connector domain shows no structural similarity to SH3 domains, as was previously suggested on the basis of sequence homology.

The SH2 domain of Stat3 shares with other SH2 domains a central three-stranded β -pleated sheet (strands B, C and D) flanked by helix αA and strands βA and βG . (Secondary-structure elements and residues are denoted according to established nomenclature¹⁸.)



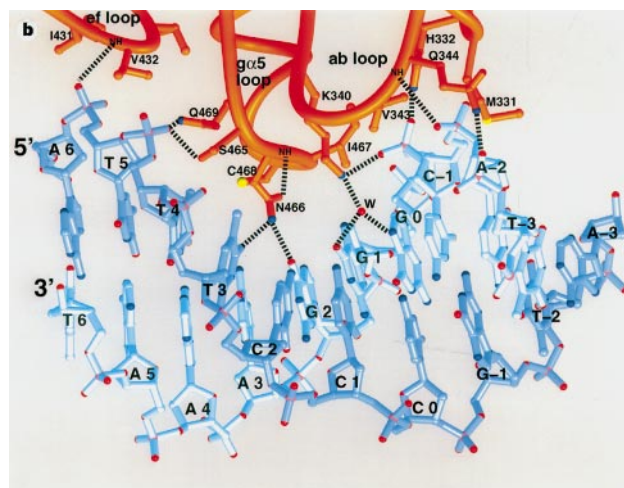
Figure 2 Ribbon diagram of the Stat3 homodimer-DNA complex. The N-terminal 4-helix bundle is shown in blue, the β -barrel domain in red, the connector domain in green, and the SH2 domain and phosphotyrosine-containing region in yellow. Disordered regions between helices $\alpha 1$ and $\alpha 2$ and residues 689 to 701

have been modelled in grey. This figure and Fig. 3b were produced with program RIBBONS⁴⁵. Views are shown **a**, along the DNA axis (the dyad of the complex running vertically); **b**, from the side, with monomer 2 depicted in grey; and **c**, from the top. Phosphotyrosines are indicated by a Y in **c**.

Structure of the DNA

DNA binding

Residues of loop $\alpha 5$ are important determinants of base specificity. In particular, Asn 466 recognizes bases in base pair ± 2 and ± 3 . The high-affinity binding site M67 site TTCCCGTAA differs at position -2 from a consensus palindromic 9-bp target site, TTCCCGGAA, identified in binding-site selection experiments for Stat1 and Stat3 (ref. 8). In the M67 site a C-G or a T-A base pair is present in position 2 and -2 , respectively. The side chain atom N $\delta 2$ of Asn 466 forms a hydrogen bond with either the exocyclic oxygen O6 of guanine (base pair 2) or O4 of thymine (base pair -2), and with the exocyclic oxygen O4 of thymine in base pairs ± 3 . At the same time, atom O $\delta 1$ of Asn 466 hydrogen-bonds with the backbone amide of residue 467, thereby anchoring Asn 466 and guaranteeing specificity. The electron density for Asn 466 is excellent, which indicates that only minor adjustments of the side chain are needed to hydrogen-bond to guanine O6 or thymine O4 of bp ± 2 . In addition, the methylene group of Asn 466 is part of a



clarity. Polar interactions are indicated by dashed lines. At non-palindromic positions (bp ± 2 , bp 0), the opposite base was built in manually, as only one DNA strand had been included into the crystallographic refinement. Note that the included water molecule shows strong density, indicating that it is present in both possible orientations of the complex around the dyad.

hydrophobic binding pocket formed by side-chain atoms of Ser 465, Asn 466, Cys 468 and Gln 469. This pocket accommodates the methyl groups of thymines in base pairs ± 3 and ± 4 , which are further critical determinants of the consensus DNA sequence.

The bases of the inner base pairs (bp + 1, bp 0 and bp - 1) are not in direct contact with the protein, in agreement with the high sequence variability of the DNA in natural target sites at these positions. However, there is a dependence of Stat1 and Stat3 DNA-binding activity on these three base pairs *in vivo*¹⁹. This might be explained by water-mediated contacts, such as that formed between Lys 340 and the bases at positions 0 and ± 1 (Fig. 3), or by 'indirect' preferences for bases which allow bending of the DNA.

All the DNA-contacting loops are well buttressed by the protein and linked to one another: strand g' (residues 462–464) and strand a' (residues 337–341) form a small β -sheet, thereby connecting loops $\alpha 5$ and ab. The carbonyl of residue 463 (loop $\alpha 5$) forms a hydrogen bond to the side chain of Arg 382 (loop cx), which in turn forms a salt bridge with Glu 435 (loop ef), thus linking these three loops as well. In the Stat3 mutant VTEEL⁴³⁶ $\rightarrow$ VTAAL, the hydrogen bond between Glu 435 and Arg 382 is disrupted, whereas in mutant SLPLVVVISN⁴⁶⁶ $\rightarrow$ SLPAAAINSN loop $\alpha 5$ becomes more flexible⁸. Thus both mutations would reduce the rigidity and buttressing of the DNA-contacting loops. In Stat6 a similar mutant SLPLVVIVH $\rightarrow$ SLPLEAAVH²⁰ introduces a charged residue into the hydrophobic core of the β -barrel.

For Stat1, Stat3, Stat4 and Stat5, a half-site spacing of three base pairs (TTCN₃GAA) is generally observed, although binding to half-sites separated by two base pairs is also possible²¹. In contrast, Stat 6 binds preferably to half-sites spaced by four base pairs²⁰. This difference could be caused by variations of specific DNA-contacting residues. For example, Ser 465 and Cys 468 in Stat3 are replaced by valine and asparagine in Stat6, respectively; Asn 466—conserved in Stat1 to Stat4 and critical for sequence-specific recognition in Stat3—is a histidine in Stat5 and Stat6. However, this histidine probably exerts a similar function as Asn 466. Therefore, it is more likely that different STAT proteins can adapt to differently spaced half-sites by reorienting their DNA-binding domains. The high thermal mobility of the SH2 domain indicates that it is not rigidly attached to the rest of the molecule, which would allow the STAT proteins to reorient their DNA-binding domains with respect to the

dimerization surface. A similar variability is seen in the crystal structures of human and mouse NF- κ B P50, which were co-crystallized with differently separated half-sites²².

Phosphotyrosine binding and dimerization

The classic phosphotyrosine peptide–SH2 domain interactions are observed in our STAT structure. Residues Lys 591 (α A2), Arg 609 (β B5), Ser 611 (β B7) and Ser 613 (β C2) form direct polar interactions with phosphotyrosine 705 and are strongly conserved in all SH2 domains²³. The side chain of Leu 706 packs against a hydrophobic pocket formed by residues of strands C and D. All STAT proteins have a hydrophobic residue in this position.

Residues 689–701 connecting the SH2 domain to the phosphotyrosine peptide (residues 702–716) are disordered. However, model building shows that these 12 residues are too short to link the SH2 domain covalently with its cognate phosphopeptide. In fact, the linker is even shorter in Stat2, Stat5 and Stat6. We conclude that the phosphotyrosine of one monomer is bound to the SH2 domain of the other. Despite poor density for residues following the phosphotyrosine, the general course of the peptide backbone can be followed (see Methods): while the phosphotyrosine and neighbouring residues (702–709) are bound *in trans* to the opposite monomer, the peptide then runs back, lines up with the dimerically related peptide over residues 708–712, and finally interacts through residues 712–716 *in cis* with the BG loop of its own SH2 domain (Figs 2c, 4). In this way, the phosphotyrosine peptides clamp the two monomers together. The side chains of residues Thr 708, Phe 710 and Cys 712 intercalate with side chains of the dimerically related peptide. Differences in the sequence of these intercalating side chains might explain the preferences for homo- or heterodimerization of different STAT family members²⁴. Residues in loop DB of the two SH2 domains also take part in the dimer interface. This region contains two Gln and two Asn residues (643–644 and 647–648) that point towards the other monomer. These interactions would not depend on phosphotyrosine binding and might explain the loose association that occurs before phosphorylation²⁵.

Interaction with other proteins

For activation, STAT proteins are recruited to the receptor through the binding of phosphotyrosine peptides by the SH2 domain. For

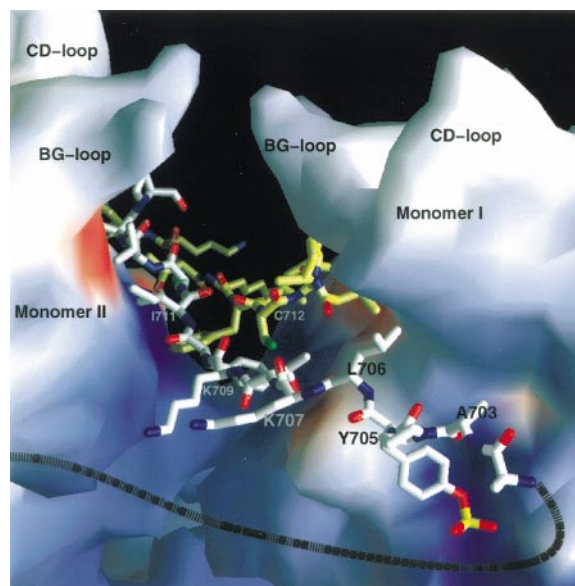


Figure 4 Binding of the phosphotyrosine peptides by the C-terminal SH2 domains. Carbon atoms of the phosphotyrosine peptide covalently linked to

monomer I and monomer II are yellow and white, respectively. Produced with program GRASP⁴⁶.

Stat3, sequences of gp130, the interleukin-6-signal transducer containing the motif YX₁X₂Q, have been defined as critical determinants for recruitment²⁶. In position X₁, Leu, Phe, Met and Arg, but not Asp, can be accepted²⁷. Leu, Phe and Met can be accommodated in the hydrophobic pocket described above, but it is not understood how arginine is bound in position X₁ and why glutamine in position Y + 3 is important.

In the nucleus, STAT proteins associate with several other factors upon DNA binding. The best-characterized example is the DNA-binding protein p48, which is bound by the Stat1/Stat2 heterodimer. It has been shown²⁸ that residues 150–250 of the heterodimer, in particular Lys 161 of Stat1, are implicated in this interaction. In our structure, these residues are part of the 4-helix bundle. The corresponding Lys 163 is part of helix α 1 and points towards the gf loop of the β -barrel (Fig. 2a). This suggests that p48 binds between the two helix bundles of the STAT heterodimer by packing against helix α 1 and the end of the β -barrel distal to the DNA, and that it contacts the DNA in the major groove.

Dimerization of the hook-like N-terminal domains via an excessive polar interface mediates cooperative binding of Stat proteins to multiple DNA target sites^{6,7,13}. The last helix of the N-terminal domain and helix H1 of our structure are separated by an 18-residue linker. Model building suggests that on tandem sites with the central base pairs separated by 20 base pairs, the two dimers are related by a translation of 70 Å along the DNA axis; therefore the distances between the N-terminal residue 136 of one monomer bound to the first site and the corresponding residues of monomers 1 and 2 bound to the second site would be 70 Å and 100 Å, respectively, which both could be spanned by the N-terminal dimers and the linkers.

Comparison with NF- κ B/Rel proteins

Despite no obvious sequence homology, we observe a striking structural and functional similarity between the STAT and Rel/NF- κ B family of transcription factors^{29–33}. Dimerization is necessary for both before they can bind DNA and the appearance of the dimers bound to DNA is similar. Furthermore, both STAT and Rel proteins contact DNA by using loops protruding from a β -barrel domain with a variant of the immunoglobulin variable fold. In both cases the β -barrels share the same topology (r.m.s._{104C α -atoms} = 3.4 Å), except for one additional strand y in NF- κ B inserted between strands x and c'. The ab loop in STAT proteins corresponds to the recognition loop in NF- κ B proteins, which is also observed in the T domain of the transcription factor *Brachyury*³⁴. STAT proteins contact the DNA through 8 loops (4 loops per monomer), whereas NF- κ B uses 10 loops (5 loops per monomer). However, the difference in binding constants, which are in the nanomolar range for STATs and in the picomolar range for NF- κ B proteins, cannot simply be explained by the difference in the size of accessible surface area buried upon DNA binding (Stat3 β :2,700 Å²; NF- κ B p52:3,200 Å²). This difference may be due to the number of direct base contacts: Stat3 β contacts only two bases per half-site, whereas NF- κ B proteins contact up to seven bases per half-site. Furthermore, the proteins are positioned differently on DNA: the dimerization domains of STAT protein and Rel proteins sit above the minor and major groove, respectively. In STAT proteins, dimerization and DNA binding is accomplished by different domains that are separated by a 'connector domain'; in Rel proteins, dimerization and DNA-contacting regions form one continuous surface. □

Methods

Protein expression, purification and crystallization. A mouse Stat3 β ¹² expression construct coding for amino-acid residues 127–722 was prepared by polymerase chain reaction (PCR) amplification and ligated into the expression vector pET32a (Novagen). The protein was expressed and tyrosine-phosphorylated in bacteria. Details of Stat3 β expression and activation will be published elsewhere. Native Stat3 β and selenomethionine-substituted

protein was purified from bacterial lysates by precipitation with polyethylenimine and ammonium sulphate, followed by gel filtration chromatography (Superose-12, Pharmacia). Chemically synthesized oligonucleotides were purified by anion-exchange chromatography (MonoQ, Pharmacia) at pH 12, dialysed against distilled water and lyophilized. Sequences of the two oligonucleotide strands are given in Fig. 3a.

For crystallization, 0.074 mM (10 mg ml⁻¹) protein dimer in 20 mM HEPES, pH 7.0, 200 mM NaCl, 10 mM MgCl₂, 5 mM DTT, 0.5 mM PMSF was mixed with an equal amount of previously annealed DNA duplex, which was added from a 0.9 mM aqueous stock solution. Crystals were grown by the hanging-drop method above a reservoir containing 0.1–0.4 M NaCl, 5 mM MgSO₄, 50 mM MES, pH 5.6–6.0, 0.1 M ammonium acetate, 10% glycerol (v/v). 2–5 μ l of protein–DNA complex solution were mixed with an equal volume of reservoir solution. Bipyramidal crystals grew within one week to an average size of 400 μ m $\times$ 400 μ m $\times$ 150 μ m.

Data collection, structure determination and refinement. Crystals were transferred stepwise to solutions of reservoir buffer containing increasing amounts of glycerol up to a maximal concentration of 35% (v/v) and frozen at 100 K in a stream of nitrogen. All data were collected at 100 K using synchrotron radiation on various beam lines of the European Synchrotron Radiation Facility (ESRF), Grenoble (Table 1). Data at ID2 and BM14 were recorded on a 345 Mar Research image plate detector and at ID14-3 on 40 cm $\times$ 80 cm image plates with an offline scanner (EMBL, Grenoble outstation). Diffraction data were integrated and scaled using programs DENZO and SCALEPACK³⁵ and further processed with programs of the CCP4 suite³⁶.

The crystals belong to spacegroup *I*₄ (*a* = *b* = 174.0 Å, *c* = 79.4 Å) and contain half a complex in the asymmetric unit. The pseudodyad of the DNA duplex coincides with the crystallographic dyad in the crystal. To exclude a lower-symmetry spacegroup as the possible reason for this observation, the data were reintegrated in subgroup *P*₄ lacking the crystallographic dyad. Simulated-annealing omit maps calculated in space group *P*₄ of a model without the DNA bases show 'averaged' density for both bases of the non-palindromic base pairs, which confirms the assigned spacegroup *I*₄ and shows that the complex is randomly oriented around the dyad.

Heavy-atom positions were located by difference-Patterson and difference-Fourier synthesis. The three mercury compounds were bound to similar sites but with different occupancies. Heavy-atom parameters were refined using programs MLPHARE³⁶ and SHARP³⁷. The resulting MIR phases were further improved by solvent flattening with program DM³⁶.

Although the electron density map calculated from experimental phases was immediately interpretable for the 4-helix bundle, the β -barrel, the connector domain and the DNA, the quality was inferior for the SH2 domain. Although the general course of the polypeptide backbone for this domain was clear, there were many breaks in the middle of strands, and side-chain density was weak or missing. The density is greatly improved in our final 2*F*_o – *F*_c map but is still noticeably poorer than in other domains. Hence, we cannot rule out errors such as a shift in register in certain loop regions and strands. Between loops DB and BG, our model accounts for Met655 and Met660 but the tracing remains ambiguous due to the poor quality of the electron density. Furthermore, two tracings of the phosphotyrosine peptide region are possible: in one tracing, the peptide (residues 708–712) forms an extended chain and runs antiparallel to the dimerically related phosphopeptide; in the other, the peptide forms a hairpin containing a tight turn which abuts with the tight turn of the dimerically related hairpin. The two tracings differ only in the position of Ile 711 and are equally consistent with the density and a mercury bound to Cys 712. However, the second tracing is unlikely as it would result in strained geometry for the isoleucine and would require the symmetry-related hairpins to align perfectly, with the strands of one collinear with those (not related by symmetry) of the other. It is well known that such a pair of interrupted, adjacent β -strands typically represents an incorrect chain tracing³⁸, and we therefore prefer the first tracing, which is shown in Figs 2 and 4. Owing to high thermal mobility, the SH2 domain and the phosphopeptide contributes very little to the scattering. We could refine a model from which the SH2 domain and the phosphotyrosine peptide was omitted to an *R*_{free} of 32.8% (*R*-factor = 27.8%).

Our model was built with program O (ref. 39) and refined with program X-PLOR⁴⁰. So far, only one DNA strand (5'-TGCAATTCCTGAAATCT-3';

Fig. 3a) was included in the crystallographic refinement. Our model has an R -factor of 25.3% ($R_{\text{free}} = 30.6\%$)⁴¹ and excellent geometry (r.m.s._{bond} = 0.13 Å, r.m.s._{angle} = 2.1 deg). The Ramachandran plot shows 78.8% of residues in the most favoured regions, 16.8% of residues in the additional allowed regions and only 6 residues in the disallowed regions, which are part of poorly ordered loops (calculated according to program PROCHECK⁴²).

Received 23 February; accepted 1 June 1998.

- Darnell, J. E. Jr STATs and gene regulation. *Science* **277**, 1630–1635 (1997).
- Ihle, J. N. STATs: Signal transducers and activators of transcription. *Cell* **84**, 331–334 (1996).
- Pellegrini, S. & Dusanter-Fourt, I. The structure, regulation and function of the Janus kinases (JAKs) and the signal transducers and activators of transcription (STATs). *Eur. J. Biochem.* **248**, 615–633 (1997).
- Kumar, A., Commune, M., Flickinger, T. W., Horvath, C. M. & Stark, G. R. Defective TNF- α -induced apoptosis in STAT1-null cells due to low constitutive levels of caspases. *Science* **278**, 1630–1632 (1997).
- Pfeffer, L. M. *et al.* Stat3 as an adaptor to couple phosphatidylinositol 3-kinase to the IFNAR1 chain of the type I interferon receptor. *Science* **276**, 1418–1420 (1997).
- Xu, X., Sun, Y.-L. & Hoey, T. Cooperative DNA binding and sequence-selective recognition conferred by the STAT amino-terminal domain. *Science* **273**, 794–797 (1996).
- Vinkemeier, U. *et al.* DNA binding of *in vitro* activated Stat1 α , Stat1 β and truncated Stat1: interaction between NH₂-terminal domains stabilizes binding of two dimers to tandem DNA sites. *EMBO J.* **15**, 5616–5626 (1996).
- Horvath, C. M., Wen, Z. & Darnell, J. E. Jr A STAT protein domain that determines DNA sequence recognition suggests a novel DNA-binding domain. *Genes Dev.* **9**, 984–994 (1995).
- Shuai, K. *et al.* Interferon activation of the transcription factor Stat91 involves dimerization through SH2-phosphotyrosyl peptide interactions. *Cell* **76**, 821–828 (1994).
- Wen, Z., Zhong, Z. & Darnell, J. E. Jr Maximal activation of transcription by Stat1 and Stat3 requires both tyrosine and serine phosphorylation. *Cell* **82**, 241–250 (1995).
- Caldenhoven, E. *et al.* Stat3 β , a splice variant of transcription factor Stat3, is a dominant negative regulator of transcription. *J. Biol. Chem.* **271**, 13221–13227 (1996).
- Schaeffer, T. S., Sanders, L. K. & Nathans, D. Cooperative transcriptional activity of Jun and Stat3 β , a short form of Stat3. *Proc. Natl Acad. Sci. USA* **92**, 9097–9101 (1995).
- Vinkemeier, U., Moarefi, I., Darnell, J. E. Jr & Kuriyan, J. Structure of the amino-terminal protein interaction domain of STAT-4. *Science* **279**, 1048–1052 (1998).
- Wagner, B. J., Hayes, T. E., Hoban, C. J. & Cochran, B. H. The SIF binding element confers *sis*/PDGF inducibility onto the *c-fos* promoter. *EMBO J.* **9**, 4477–4484 (1990).
- Fu, X.-Y., Schindler, C., Improta, T., Aebersold, R. & Darnell, J. E. Jr The proteins of ISGF-3, the interferon α -induced transcriptional activator, define a gene family involved in signal transduction. *Proc. Natl Acad. Sci. USA* **89**, 7840–7843 (1992).
- Herzberg, O. & James, M. N. G. Structure of the calcium regulatory muscle protein troponin C at 2.8 Å resolution. *Nature* **313**, 653–659 (1985).
- Sundaralingam, M., Bergstrom, R., Strasburg, G. & Rao, S. T. Molecular structure of troponin C from chicken skeletal muscle at 3 angstrom resolution. *Science* **227**, 945–948 (1985).
- Eck, M. J., Shoelson, S. E. & Harrison, S. C. Recognition of a high-affinity phosphotyrosyl peptide by the Src homology-2 domain of p56^{lck}. *Nature* **362**, 87–91 (1993).
- Lamb, P. *et al.* STAT protein complexes activated by interferon- α and gp130 signalling molecules differ in their sequence preferences and transcriptional induction properties. *Nucleic Acids Res.* **23**, 3283–3289 (1995).
- Mikita, T., Campbell, D., Wu, P., Williamson, K. & Schindler, U. Requirements for interleukin-4-induced gene expression and functional characterization of Stat6. *Mol. Cell. Biol.* **16**, 5811–5821 (1996).
- Seidel, H. M. *et al.* Spacing of palindromic half sites as a determinant of selective STAT (signal transducers and activators of transcription) DNA and transcriptional activity. *Proc. Natl Acad. Sci. USA* **92**, 3041–3045 (1995).
- Müller, C. W., Rey, F. A. & Harrison, S. C. Comparison of two different DNA-binding modes of the NF- κ B p50 homodimer. *Nature Struct. Biol.* **3**, 224–227 (1996).
- Kuriyan, J. & Crothers, D. Modular peptide recognition domains in eukaryotic signaling. *Annu. Rev. Biophys. Biomol. Struct.* **26**, 259–288 (1997).
- Heim, M. H., Kerr, I. M., Stark, G. R. & Darnell, J. E. Jr Contribution of Stat SH2 groups to specific interferon signaling by the Jak-Stat pathway. *Science* **267**, 1347–1349 (1995).
- Stancato, L. F., David, M., Carter-Su, C. & Lerner, A. C. Preassociation of Stat1 with Stat2 and Stat3 in separate signalling complexes prior to cytokine stimulation. *J. Biol. Chem.* **271**, 4134–4137 (1996).
- Stahl, N. *et al.* Choice of STATs and other substrates specified by modular tyrosine-based motifs in cytokine receptors. *Science* **267**, 1349–1353 (1995).
- Gerhart, C. *et al.* Differential activation of acute phase response factor/Stat3 and Stat1 via the cytoplasmic domain of the interleukin 6 signal transducer gp130. *J. Biol. Chem.* **271**, 12991–12998 (1996).
- Horvath, C. M., Stark, G. R., Kerr, I. M. & Darnell, J. E. Jr Interactions between STAT and non-STAT proteins in the interferon-stimulated gene factor 3 transcription complex. *Mol. Cell. Biol.* **16**, 6957–6964 (1996).
- Müller, C. W., Rey, F. A., Sodeoka, M., Verdine, G. L. & Harrison, S. C. Structure of the NF- κ B p50 homodimer bound to DNA. *Nature* **373**, 311–317 (1995).
- Ghosh, G., Van Duyn, G., Ghosh, S. & Sigler, P. B. The structure of NF- κ B p50 homodimer bound to a κ B site. *Nature* **373**, 303–310 (1995).
- Cramer, P., Larson, C. J., Verdine, G. L. & Müller, C. W. Structure of the human NF- κ B p52 homodimer–DNA complex at 2.1 Å resolution. *EMBO J.* **16**, 7078–7090 (1997).
- Chen, Y.-Q., Ghosh, S. & Ghosh, G. A novel DNA recognition mode by the NF- κ B p65 homodimer. *Nature Struct. Biol.* **5**, 67–73 (1998).
- Chen, F. E., Huang, D.-B., Chen, Y.-Q. & Ghosh, G. Crystal structure of p50/p65 heterodimer of transcription factor NF- κ B bound to DNA. *Nature* **391**, 410–413 (1998).
- Müller, C. W. & Herrmann, B. Crystallographic structure of the T domain–DNA complex of the Brachyury transcription factor. *Nature* **389**, 880–884 (1997).
- Otwinski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Meth. Enzymol.* **276**, 307–326 (1997).
- Collaborative Computational Project Number 4. The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–776 (1994).
- Fortelle, E. D. L. & Brice, G. Maximum-likelihood heavy-atom parameter refinement for the multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Meth. Enzymol.* **276**, 472–494 (1997).
- Richardson, J. S. & Richardson, D. C. Interpretation of electron density. *Meth. Enzymol.* **115**, 189–206 (1985).
- Jones, T. A., Zhou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Brünger, A. T. *X-PLOR version 3.1. A system for X-ray crystallography and NMR* (Yale University Press, New Haven, CT, 1992).
- Brünger, A. T. Free R value: A novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–475 (1992).
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* **26**, 283–291 (1993).
- Higgins, D. G., Thompson, J. D. & Gibson, T. J. Using CLUSTAL for multiple sequence alignments. *Meth. Enzymol.* **266**, 383–402 (1996).
- Chen, L., Glover, J. N. M., Hogan, P. G., Rao, A. & Harrison, S. C. Structure of the DNA-binding domains from NFAT, Fos and Jun bound specifically to DNA. *Nature* **392**, 42–48 (1998).
- Carson, M. Ribbons 2.0. *J. Appl. Crystallogr.* **24**, 958–961 (1991).
- Nichols, A., Bharadwaj, R. & Honig, B. GRASP—graphical representation and analysis of surface properties. *Biophysics J.* **64**, A166–A166 (1993).

Acknowledgements. We thank our colleagues for their help and for discussions, particularly C. Petosa for help with refinement, A. Perrakis for help with program Sharp, and C. Petosa, F. Baudin, J. Grimes and P. Cramer for comments on the manuscript; we also thank members of the EMBL/ESRF Joint Structural Biology Group, particularly B. Rasmussen, J. Grimes and J. Lescar at ID2, A. Thompson, V. Stojanoff and G. Leonard at BM14, and W. Burmeister and S. Wakatsuki at ID14-3, for their support; and R. Eritja for oligonucleotide synthesis. This project was supported by an institutional fellowship of the European Union and a short-term EMBO fellowship to S.B.

Correspondence and requests for materials should be addressed to C.W.M. (e-mail: mueller@embl-grenoble.fr). Atomic coordinates have been deposited with the Protein Data Bank (Brookhaven National Laboratory) under accession number 1bg1.

YOURS TO HAVE AND TO HOLD BUT NOT TO COPY

The publication you are reading is protected by copyright law.
Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA.
Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited
90 Tottenham Court Road, London W1P 0LP
Telephone: 0171 436 5931 Fax: 0171 436 3986

Large-scale coronal heating by the small-scale magnetic field of the Sun

C. J. Schrijver*, A. M. Title*, K. L. Harvey†, N. R. Sheeley Jr‡, Y.-M. Wang‡, G. H. J. van den Oord§, R. A. Shine*, T. D. Tarbell* & N. E. Hurlburt*

* Stanford-Lockheed Institute for Solar and Astrophysics (H1-12/252), 3251 Hanover Street, Palo Alto, California 94034, USA

† Solar Physics Research Corporation, 4720 Calle Desecada, Tucson, Arizona 85718, USA

‡ Code 7672, E. O. Hulburt Center for Space Research, Naval Research Laboratory, Washington DC 20375-5352, USA

§ Astronomical Institute, University of Utrecht, PO Box 80,000, 3508 TA Utrecht, The Netherlands

Magnetic fields play a crucial role in heating the outer atmospheres of the Sun and Sun-like stars, but the mechanisms by which magnetic energy in the photosphere is converted to thermal energy in the corona remain unclear. Observations show that magnetic fields emerge onto the solar surface as bipolar regions with a broad range of length scales. On large scales, the bipolar regions survive for months before dispersing diffusively^{1–3}. On the smaller scales, individual bipolar regions disappear within days but are continuously replenished by new small flux concentrations, resulting in a sustained state of mixed polarity⁴. Here we determine the rate of emergence of these small bipolar regions and we argue that the frequent magnetic reconnections associated with these regions (an unavoidable consequence of continued flux replacement) will heat the solar atmosphere. The model that describes the details of these mixed-polarity regions⁴ is complementary to the traditional diffusion model for large-scale flux dispersal^{1–3}, and a combination of the two should lead to a more complete understanding of the role of magnetic fields in stellar atmospheres.

Magnetic field emerges onto the visible solar surface—the photosphere—in bipolar regions that range in flux from below the detection threshold of somewhat more than $\sim 10^{16}$ maxwells (10^8 Mx equals 1 Wb) up to $\sim 5 \times 10^{22}$ Mx. Those regions whose flux exceeds $\sim 3 \times 10^{20}$ Mx are called active regions; these frequently contain sunspots early in their development. In the quiet regions surrounding them, the flux is located in a multitude of small concentrations (Fig. 1) that are positioned within, and evolve with, the downflow regions in the supergranular convection, which form a network-like pattern with cell sizes that range from $\sim 10,000$ km to $\sim 30,000$ km (ref. 5). Much of this flux originates in small bipoles, the so-called ephemeral regions.

Ground-based data are subject to atmospheric distortions and frequent interruptions, so that recognizing flux emergence in ephemeral regions is difficult. The new observational data obtained with the Michelson Doppler Imager, MDI⁶, on board the Solar and Heliospheric Observatory, SOHO, are free from atmospheric distortions and are uninterrupted for long periods with a cadence of one minute, which allows us to measure important processes with highly improved accuracy. The MDI can provide magnetic maps of a region of 600×600 arcsec (1 arcsec ≈ 725 km) with an angular resolution of 1.2 arcsec. We have analysed quiet-Sun sequences spanning 8 to 22 h with a 1-min cadence. We find that ephemeral regions, with fluxes ranging from 3×10^{18} Mx to 3×10^{19} Mx, emerge rapidly. The initial emergence phase lasts ~ 30 min, during which the poles separate up to $\sim 7,000$ km, moving at ~ 4 km s⁻¹. After this initial phase, the flux drifts with the supergranular flow, slowing down by a factor of about ten.

We measure an average emergence rate of ephemeral regions near disk centre equivalent to ~ 1 per area of $27,000$ km² per day. The mean total absolute flux per ephemeral region is 1.3×10^{19} Mx, in good agreement with recent ground-based estimates⁴. With this rate of emergence, it takes only ~ 40 h for as much flux to surface in the quiet photosphere as is present there. Flux does not build up, however, because it disappears again in collisions involving opposite polarities. Schrijver *et al.*⁴ developed a statistical description of flux collisions and fragmentations to model the distribution of fluxes in concentrations. They model flux concentrations as short-lived structures that are deformed, broken up, cluster in collisions involving like polarities, or cancel in collisions against opposite polarities. They find that the timescale for all of the flux concentrations in the quiet photosphere to cancel against others lies between 1.5 and 3 d, consistent with the observed emergence rate.

Within a flux replacement timescale $\hat{t}$ of 40 h, the distance over which flux can travel in its characteristic random walk before being cancelled is of the order of $l \approx 2 \sqrt{D_s \hat{t}}$ (with $D_s \approx 300$ km² s⁻¹ (ref. 7) the diffusion coefficient for medium-term flux dispersal), or 13,000 km. Hence, in order to ensure the observed continued small-scale mixing of polarities in quiet regions, ephemeral regions should emerge with a statistically homogeneous probability on length scales as low as a single supergranule and an associated timescale of a few days, compatible with what is observed.

The solar surface also contains the intranetwork field that emerges in patches with a flux distribution that peaks near 5×10^{16} Mx. The effect of this intranetwork flux on the longer-lived, much larger concentrations discussed above appears small: in very quiet regions, at least 90% of the network concentrations originate in ephemeral regions, while the remaining 10% or less is caused by the piling up of intranetwork patches^{8,9}. The short lifetime of intranetwork concentrations and their very small size are likely to limit their direct effect on the outer solar atmosphere.

Owing to the fine-scale mixing of the polarities on the solar surface, the quiet corona (the outermost solar atmosphere with temperatures of $>10^6$ K) has a very complicated field geometry, with interlaced connections on a range of length scales. The random insertion of small bipolar regions and the subsequent disappearance of newly paired, previously unconnected concentrations inevitably forces continual reconnection in the coronal field. These reconnections result in atmospheric heating. The compact coronal bright points¹⁰ stand out in images of the quiet corona, but much of the emission originates from an unresolved background haze over the ensemble of flux concentrations.

Among the proposed mechanisms to heat the corona over the quiet photosphere, one category relies on wave dissipation, the other on current dissipation. The latter includes “braiding” of the coronal field lines that occurs because of the random shuffling of footpoints caught in the photospheric turbulence^{11–13}. If we consider only this braiding, the rate at which energy is supplied to the corona is given by a Poynting flux of $F \propto B_0^2 v^2 t/L$, for an average photospheric flux density B_0 , mean displacement velocity v , correlation timescale t , and loop length L (refs 11, 13).

The dependence of the Poynting flux on loop length allows efficient heating in the many short connections between nearby concentrations of opposite polarity; this heating occurs by the process of footpoint braiding by means of granular convection with a scale of 1,000 km (refs 11, 13). We propose that small-scale flux replacement results in efficient heating of the longer, higher-arching connections, by braiding of the large-scale coronal field. That is, whenever flux from a newly emerged small bipole cancels against existing flux of opposite polarity, part of the flux of that opposite polarity flux is in effect ‘moved’ to the location of the other pole of the bipole. The cancelled flux has been replaced by the surviving opposite-polarity flux at another location (not to be interpreted as diffusive dispersal¹⁴). This footpoint reconnection leads to braiding, because one footpoint is being effectively moved

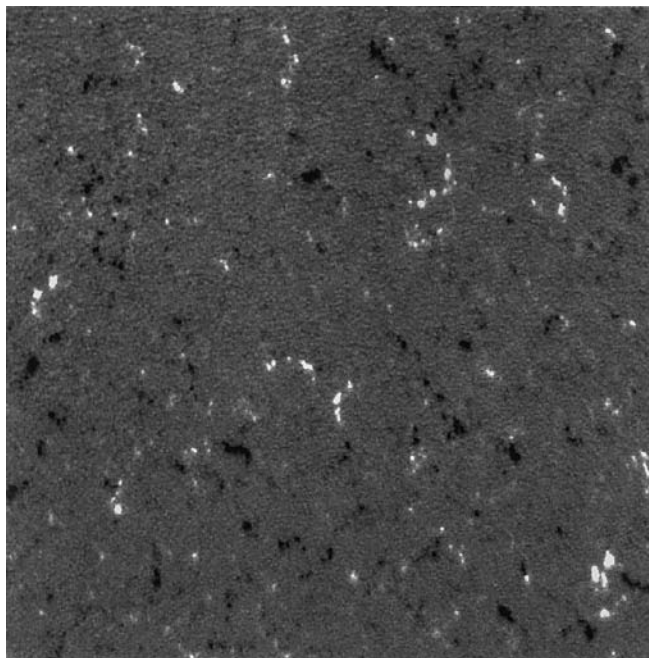


Figure 1 Magnetic map of a quiet part of the solar disk ($130,000 \times 130,000$ km). No substantial active regions emerged in this region for more than a month. Flux disappears in collisions between opposite polarities (here shown black and

white) so fast that all field should disappear in a few days. New flux emerging in small ephemeral regions replaces the disappearing flux, resulting in this thoroughly mixed pattern.

over a distance small compared to the loop length. The associated part of the Poynting flux for a given B_0^2/L then scales with $v^2 t$. For braiding on the scale of granular convection (1,000 km), $v^2 t$ can be derived from the observed short-term r.m.s. displacements of flux concentrations which can be characterized by a diffusion coefficient D_g : $v^2 \hat{t} = 4D_g \approx 300 \text{ km}^2 \text{ s}^{-1}$ (ref. 15). The effective braiding associated with flux replacement involves a step length across a supergranule of 15,000 to 18,000 km on a timescale $\hat{t}$ of 4 to 6 h. These braiding events occur relatively infrequently, because the total flux replacement timescale is some 40 h. Correcting for this by the ratio of $\hat{t}$ to flux replacement time, the effective product $v^2 \hat{t}$ reaches $v^2 \hat{t} \times \hat{t}/40 \approx 2,000 \text{ km}^2 \text{ s}^{-1}$, which is much larger than that associated with granular braiding.

Apart from contributing to coronal heating, the 'clipping' of coronal field lines (caused by the rapid flux replacement in the quiet photosphere) is likely to act as the 'lubricant' by which the coronal fields relax the shear that is built up by the differential rotation. Reconnection would eventually occur regardless of the mechanism, but flux replacement appears to be an efficient way to make it happen even if the stresses are still small. Future studies will need to address whether it is in fact the most important mechanism. The nearly homogeneous emergence of essentially randomly orientated ephemeral regions does not, of course, by itself lead to relaxation of stresses. Instead, the relaxation is a consequence of the braided state of the field: whenever reconnection takes place, the field is likely to relax to a smoother, less-energetic state with smaller currents.

When viewed on scales larger than some tens of thousands of kilometres, the large-scale patterns in the photospheric magnetic field are described remarkably well by a surface diffusion model¹ in which the field spreads diffusively, passively responding to the differential rotation and a slow, poleward meridional flow^{2,3}. This diffusion is the result of the supergranular flows that force the flux to perform a random-walk motion. But at the coarse angular resolution on which this model applies, much of the total flux is missed, because the mixed-polarity field results in numerical cancellation of flux of opposite polarities within resolution elements. The difference can be substantial. In the inactive phase of the 1-yr solar activity cycle, such as during the early part of 1997, most of the solar surface was covered by a mixed-polarity field in which

fluxes in the two polarities balanced to ~ 0.1 G at a resolution of $\sim 50,000$ km, whereas the average absolute flux density at the spatial resolution of ~ 900 km, was ~ 2 G; that is, an order of magnitude greater. This big difference is important for our understanding of the solar dynamo process.

Quantitative understanding of the effects of polarity-mixing on small scales is also important to, for example, the modelling of the heating of stellar outer atmospheres. The heating of the chromosphere (a domain with temperatures ranging from 6,000 to 20,000 K) of the Sun and Sun-like stars appears to depend on the total absolute flux rather than on the net flux¹⁶. As a result, extended mixed-polarity regions give rise to substantial contributions to the total emission of stars¹⁷, despite their very low flux densities on scales of the supergranulation and above. We propose that the evolution of the photospheric magnetic field can be described by two complementary models that can be applied on different length scales. Together they form the basis for detailed modelling of stellar magnetic fields and outer atmospheres. □

Received 27 August 1997; accepted 17 February 1998.

1. Leighton, R. B. Transport of magnetic fields on the sun. *Astrophys. J.* **140**, 1547–1562 (1964).
2. DeVore, C. R., Sheeley, N. R. Jr, Boris, J. P., Young, T. R. & Harvey, K. L. Numerical simulations of large-scale solar magnetic fields. *Astr. J. Phys.* **38**, 999–1007 (1985).
3. Wang, Y.-M., Nash, A. G. & Sheeley, N. R. Jr Magnetic flux transport on the sun. *Science* **245**, 712–718 (1989).
4. Schrijver, C. J., Title, A. M., van Ballegoijen, A. A., Hagenaar, H. J. & Shine, R. A. Sustaining the quiet chromospheric network: a dynamic balance of flux emergence, fragmentation, merging, and cancellation. *Astrophys. J.* **487**, 424–437 (1997).
5. Spruit, H. C., Nordlund, Å. & Title, A. M. Solar convection. *Annu. Rev. Astron. Astrophys.* **28**, 263–301 (1990).
6. Scherrer, P. H. et al. The solar oscillations investigation—michelson doppler imager. *Solar Phys.* **162**, 129–188 (1995).
7. Schrijver, C. J. et al. Dynamics of the chromospheric network: mobility, dispersal, and diffusion coefficients. *Astrophys. J.* **468**, 921–932 (1996).
8. Martin, S. F. in *Solar Photosphere: Structure, Convection, and Magnetic Fields* (ed. Stenflo, J. O.) 129–146 (IAU Symp. 138, Kluwer Academic, Dordrecht, 1990).
9. Wang, H. & Zirin, H. The velocity pattern of weak solar magnetic fields. *Solar Phys.* **115**, 205–220 (1988).
10. Harvey, K. L. The relationship between coronal bright points as seen in He I $\lambda 10830$ and the evolution of the photospheric network magnetic fields. *Astr. J. Phys.* **38**, 875–883 (1985).
11. Parker, E. N. *Spontaneous Current Sheets in Magnetic Fields* (Oxford Univ. Press, 1994).
12. Galsgaard, K. & Nordlund, Å. Heating and activity of the solar corona 1. Boundary shearing of an initially homogeneous magnetic field. *J. Geophys. Res.* **101**, 13445–13460 (1996).
13. Van Ballegoijen, A. A. Electric currents in the solar corona and the existence of magnetostatic equilibrium. *Astrophys. J.* **298**, 421–430 (1985).
14. Wang, Y.-M. & Sheeley, N. R. Jr Magnetic flux transport and the sun's dipole moment: New twists to the Babcock-Leighton model. *Astrophys. J.* **375**, 761–770 (1991).

15. Hagenaar, H. J., Schrijver, C. J., Title, A. M. & Shine, R. A. Dispersal of magnetic flux in the quiet solar photosphere. *Astrophys. J.* (submitted).
16. Schrijver, C. J., Côté, J., Zwaan, C. & Saar, S. H. Relations between the photospheric magnetic field and the emission from the outer atmospheres of cool stars. I. The solar Ca II K line core emission. *Astrophys. J.* **337**, 964–976 (1989).
17. Schrijver, C. J. in *Stellar Surface Structure* (eds Strassmeier, K. G. & Linsky, J. L.) (IAU Symp. 176, Kluwer Academic, Dordrecht, 1996).

Correspondence and requests for materials should be addressed to C.J.S. (e-mail: kschrijver@solar.stanford.edu).

The signature of chemical valence in the electrical conduction through a single-atom contact

Elke Scheer^{*†}, Nicolás Agraït[‡], Juan Carlos Cuevas[§], Alfredo Levy Yeyati[§], Bas Ludoph^{||}, Alvaro Martín-Rodero[§], Gabino Rubio Bollinger[‡], Jan M. van Ruitenbeek^{||} & Cristián Urbina^{*}

^{*} Service de Physique de l'Etat Condensé, CEA-Saclay, F-91191 Gif-sur-Yvette, France

[‡] Laboratorio de Bajas Temperaturas, Departamento de Física de la Materia Condensada C-III, Instituto Universitario de Ciencia de Materiales “Nicolás Cabrera”, Universidad Autónoma de Madrid, E-28049 Madrid, Spain

[§] Departamento de Física Teórica de la Materia Condensada C-V, Universidad Autónoma de Madrid, E-28049 Madrid, Spain

^{||} Kamerlingh Onnes Laboratorium, Leiden University, PO Box 9506, 2300 RA Leiden, The Netherlands

Fabrication of structures at the atomic scale is now possible using state-of-the-art techniques for manipulating individual atoms¹, and it may become possible to design electrical circuits atom by atom. A prerequisite for successful design is a knowledge of the relationship between the macroscopic electrical characteristics of such circuits and the quantum properties of the individual atoms used as building blocks. As a first step, we show here that the chemical valence determines the conduction properties of the simplest imaginable circuit—a one-atom contact between two metallic banks. The extended quantum states that carry the current from one bank to the other necessarily proceed through the valence orbitals of the constriction atom. It thus seems reasonable to conjecture that the number of current-carrying modes (or ‘channels’) of a one-atom contact is determined by the number of available valence orbitals, and so should strongly differ for metallic elements in different series of the periodic table. We have tested this conjecture using scanning tunnelling microscopy and mechanically controllable break-junction techniques^{2,3} to obtain atomic-size constrictions for four different metallic elements (Pb, Al, Nb and Au), covering a broad range of valences and orbital structures. Our results demonstrate unambiguously a direct link between valence orbitals and the number of conduction channels in one-atom contacts.

The electrical conductance of a quantum coherent structure accommodating N channels is given by the Landauer formula $G = G_0 \sum_{n=1}^N T_n$, where $G_0 = 2e^2/h$ is the conductance quantum (e is the electron's charge and h is Planck's constant)⁴, and T_n is the transmission probability of the n th channel. Measuring the conductance is equivalent to measuring the sum $\sum_{n=1}^N T_n$, and gives no information about the individual T_n s. A recently developed technique based on the nonlinearities in the current–voltage characteristics (referred to here as IVs) of superconducting constrictions provides the latter information⁵. These nonlinearities, called ‘sub-gap structure’⁶, are sensitive to the exact values of the individual

transmissions and can therefore be used as a tool to determine the channel ensemble, $\{T_n\}$.

In the ground state of a superconductor, all electrons are paired and there is an energy gap Δ for the creation of a quasiparticle excitation, that is, an unpaired electron. When transferring an electron between two banks in their ground state, two quasiparticles are created, one on each bank, and the energy 2Δ has to be provided by the voltage source in the circuit. The consequence is that a gap occurs in the single-electron transport characteristic for voltages $V < 2\Delta/e$. In addition, multiple Andreev reflections⁶ are possible, involving the transfer of m electrons while two quasiparticles are created. These processes, which have a probability proportional to $(T_n)^m$ for small T_n , generate current steps at voltages $V = 2\Delta/me$ within the gap. This is the origin of the sub-gap structure, the exact shape of which depends on $\{T_n\}$. Several authors^{7–9} have calculated the current voltage characteristic $i(T, V)$ of a single superconducting channel with arbitrary T .

We have used various experimental methods to achieve stable atomic-size contacts: a scanning tunnelling microscope (STM) for the measurements on Pb (ref. 10), mechanically controllable break junctions for Nb (ref. 2), and lithographically fabricated versions of the latter (ref. 3) for Al and Au. Figure 1 shows the typical development of the conductance G for the four metals as the contacts are stretched. The total conductance decreases by steps until the contact breaks. After that it decreases exponentially with distance, the signature of the tunnel regime. The step heights are in the order of G_0 , while the series of ‘plateaux’ is different for each individual opening of the contact. It has been shown that the origin of the steps is in atomic rearrangements^{11–14} which occur while the constriction is elongated. The typical value of G at which the contact breaks, the lengths of the plateaux, and the behaviour within the plateaux are characteristic of each material. Pb shows plateaux with negative slope, and an almost continuous decrease of G between $3 G_0$ and $1 G_0$, where the contact usually breaks. Al has plateaux with positive slope, and breaks at around G_0 with a jump to $G < 0.1 G_0$. Nb displays negatively inclined plateaux, smallest contacts with conductances between $2 G_0$ and $3 G_0$, and a jump to a tunnelling conductance as high as $G \approx G_0$. For our lithographic Au samples we obtain less well defined plateaux. The last contact before break typically has a conductance between $0.3 G_0$ and $0.8 G_0$.

We can stop the elongation at any point and record the IVs. With the junction in the vacuum tunnel regime, the IV corresponds well to the one expected for a single channel as shown for Pb in curve e of Fig. 2. In the contact regime one observes IVs as depicted in curves a to d of the same figure. These curves were recorded on the same sample for different last plateaux before breaking. It is usually assumed that these smallest stable contacts are one-atom contacts². Our present analysis will give further support to this assumption. The occurrence of IVs with different sub-gap structure for the same total conductance (see Fig. 2) reveals the importance of the geometrical environment of the central atom for the conduction properties. In general, the IVs in the contact regime cannot be described by the single channel theory. We shall assume that the total current can be decomposed as:

$$I(V) = \sum_{n=1}^N i(T_n, V) \quad (1)$$

where $i(T_n, V)$ is the current of channel n . (This approximation is justified as the multiple Andreev processes do not mix the normal conduction channels, see, for instance, ref. 15.) The solid curves in Fig. 2 are fits to the experimental IVs using equation (1), and the theoretical curves $i(T, V)$ calculated in ref. 8, with the set $\{T_n\}$ and the number of channels N as fitting parameters. Details of the analysis are given in ref. 5.

For Pb it turns out that three or four channels contribute to the total current for the smallest contacts. Despite the fact that more

[†] Permanent address: Physikalisches Institut, Universität Karlsruhe, D-76128 Karlsruhe, Germany.

than two channels are available, G can be well below $2 G_0$. For larger contacts with $G \geq 3 G_0$, we find six or more channels. The numbers N of channels we found at the different plateaux are indicated in Fig. 1. Only channels with a transmission larger than 1% of the total transmission are taken into account when determining N . Applying a similar analysis to different metals, we arrive at the important conclusion that the maximum number of channels $N_{\max}$ for the smallest contacts is characteristic for a given metal. In the IVs of Al

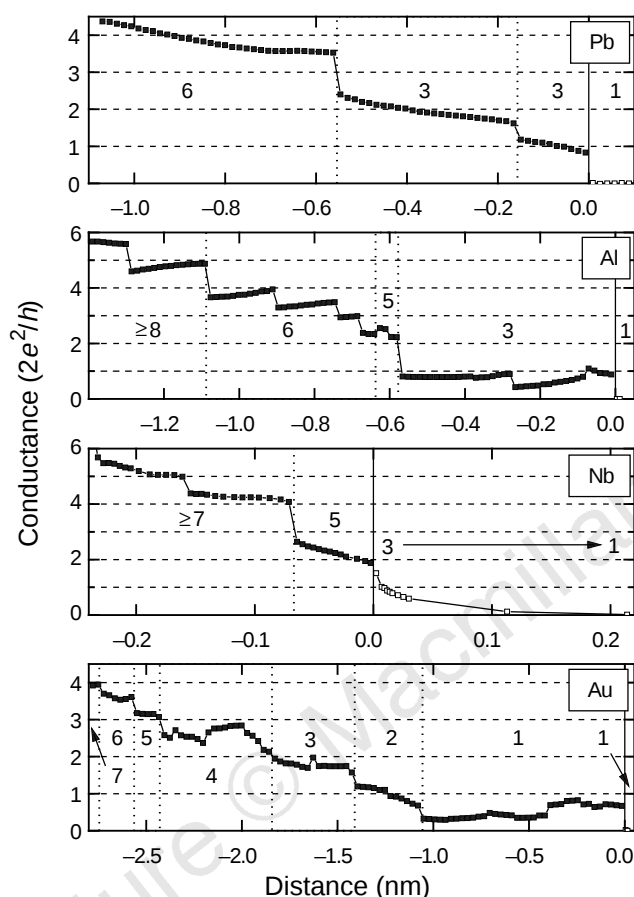


Figure 1 Typical conductance G as a function of distance, recorded during a continuous opening of the samples, for four different metals. The distribution of transmission values $\{T_n\}$ is established by fitting the IVs recorded at each point. The Pb data have been taken with a STM at 1.5 K. For preparing few-atom contacts with the STM, the Pb tip is indented repeatedly into the Pb sample surface and finally withdrawn¹⁰. For Al and Au we used lithographic break junctions below 100 mK. The initial samples before breaking are suspended nanobridges made of evaporated thin films with a central constrictions of about $200 \text{ nm} \times 200 \text{ nm}$ (ref. 3). The Nb data have been measured on a classical break junction, that is, a notched wire glued on top of a flexible substrate², at 1.6 K. By bending the substrate the break junctions are elongated and finally broken. The opening speeds were in all cases in the range $1\text{--}10 \text{ pm s}^{-1}$. The numbers indicate the number N of conduction channels in each region. The symbol " $\geq$ " means that only a lower bound for N could be determined, because of systematic deviations from the theory due to a modified quasiparticle density of states (Nb) or because of experimental scattering (Al) which limits the maximum N that can be disentangled. Plateaux corresponding to different N are separated by dotted vertical lines. The horizontal arrow indicates that N decreases in the tunnel regime. Filled symbols indicate contact regime, open symbols the tunnel regime. The origins of the distance axes are arbitrarily set at the points where the contacts break. The distance scales for Al, Nb, and Au have been extracted from the exponential distance dependence of the current in the tunnel regime. The distance scale for the measurements on Pb has been determined using the calibration of the STM and is consistent with that deduced from the tunnel regime.

smallest contacts we find typically the contributions of three channels, whereas for Nb we detect five channels. These results support our initial hypothesis that $N_{\max}$ for one-atom contacts is limited by the number of valence orbitals of the central atom N_{orb} .

To gain insight into how the conduction channels are actually built up from the atomic orbital structure we have constructed a model of a one-atom constriction using an atomic orbital basis^{16,17}. In references 16 and 17 it is explained in detail how the theoretical $\{T_n\}$ can be obtained, and how the different atomic orbitals at the central atom contribute to each conduction channel as a function of the atomic arrangement of the contact. Of course, the detailed atomic positions of the atoms in the experiments are unknown and quantitative comparison with theory is difficult. However, it turns out that the basic predictions of the model, in particular the number N of conduction channels, are independent of the exact geometry chosen for the calculation. Figure 3 illustrates schematically the main results of our theoretical analysis for a one-atom contact between two closed packed pyramids (Fig. 3a) for Al (left) and Pb (right), for which one s orbital and three p orbitals have to be considered: due to the coupling of the central atom with its neighbours, the isolated atom's energy levels (thick vertical bars in Fig. 3b and c) broaden to form a continuous local density of states (LDOS) in the neck (Fig. 3d and e). The central atom has less neighbours than an atom in the bulk material, and therefore the bands are narrower than deep inside the perfect crystal (Fig. 3b and c). In the neck geometry, the central-atom p_z orbital (z is the transport direction) splits off from the p_x and p_y orbitals and hybridizes with the s orbital, building a channel of mixed sp_z character. The actual transport properties of the atomic contact are determined by the T_n values at the Fermi level E_F which we determined self-consistently by imposing local charge neutrality (Fig. 3f and g)^{16,17}.

For the perfectly ordered Pb contact the main predictions are: (1) three channels have a significant T_n around the Fermi energy, the fourth one having $T_n < 10^{-3}$ (not visible in Fig. 3); (2) the total transmission is close to 2.5 around E_F ; (3) there is one mode which is widely open for a broad energy range, and (4) the second transmission eigenvalue is two-fold degenerate and has a smaller value

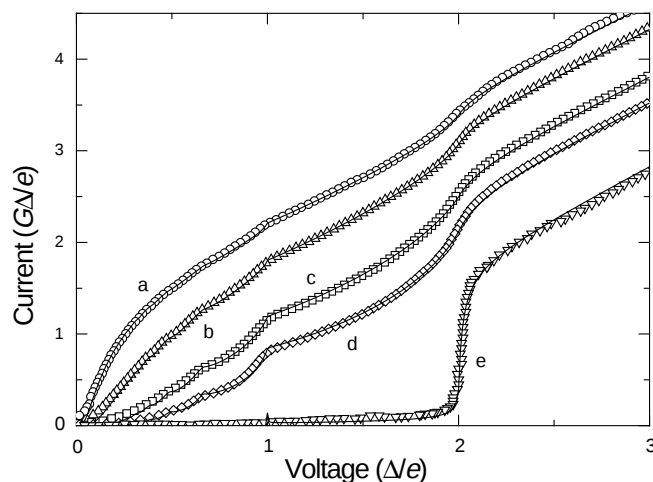


Figure 2 Measured current-voltage characteristics (plotting symbols) of five different configurations of a Pb sample at 1.5 K using STM, and best numerical fits (lines). Curves a–d have a very similar total transmission, close to 1.4; curve e has been taken in the tunnel regime. The individual channel transmissions obtained from the fits are: curve a, $T_1 = 0.955$, $T_2 = 0.355$, $T_3 = 0.085$, $T_4 = 0.005$; curve b, $T_1 = 0.89$, $T_2 = 0.36$, $T_3 = 0.145$, $T_4 = 0.005$; curve c, $T_1 = 0.76$, $T_2 = 0.34$, $T_3 = 0.27$, $T_4 = 0.02$; curve d, $T_1 = 0.65$, $T_2 = 0.34$, $T_3 = 0.29$, $T_4 = 0.12$; curve e, $T = 0.026$. Voltage and current are in reduced units. The measured superconducting gap was $\Delta/e = 1.37 \text{ mV}$.

around E_F . The widely open channel can be identified with the symmetric combination of s and p_z . The two degenerate modes mainly correspond to the p orbitals perpendicular to the z direction. The degeneracy of the p_x, p_y modes is due to the symmetry of the perfectly ordered model geometry and is lifted when disorder is introduced. Furthermore, disorder slightly reduces the T_n of all channels without altering the gross features¹⁶.

These predictions are in good agreement with our experimental observations: smallest Pb contacts just before breaking most frequently have a G between $1 G_0$ and $3 G_0$. The decomposition usually results in three or four channels, where the T_n of the fourth one, if present, is usually smaller than 0.03. We do not find last plateaux before breaking with only one or two, or with five, channels. Typically the analysis of the last plateau yields one well transmitted mode with $T_n > 0.6$ and two smaller non-degenerate ones with $T_n \leq 0.4$. We attribute the lack of degeneracy and the smaller T_n s to an asymmetrical environment around the central atom in the experimental contacts.

The case of Al is qualitatively similar because the relevant orbitals are of the same nature. However, the energy difference between the atomic s and p levels is smaller than in Pb. Furthermore, there are only three valence electrons and therefore E_F is lower. Our model predicts again three open channels for a perfect neck-shaped one-atom contact with a total conductance close to G_0 . Experimentally,

we obtain a last plateau with $G_0 \approx G_0$, decomposable into three channels⁵. The case of an Al atom sandwiched between two flat surfaces has already been treated with density functional¹⁸ and other *ab initio* calculations¹⁹, which predicted $G \approx 2 G_0$.

For Nb one-atom contacts, $N_{\max} = 6$ as the density of states around E_F is mainly determined by the contributions of one $5s$ and five $4d$ orbitals. The model calculations for Nb yield five channels with non-vanishing T_n (ref. 16) adding up to $G \approx 2.8 G_0$, again in agreement with the experimental findings (see third panel of Fig. 1). The experimental determination of $\{T_n\}$ is less precise than for Pb and Al, because the quasiparticle density of states for Nb differs slightly from the one used for calculating the $i(T, V)$. (Deviations of the quasiparticle density of states from the case considered in the theory mostly affect the shape of the current steps around $V = 2\Delta/me$ but not the current values in between. Therefore the analysis with the unmodified density of states is still possible.) An additional difference from the Pb and Al case is observed in the tunnel regime: for Nb contacts just after breaking, three channels contribute to the tunnel current. With increasing distance between the electrodes, one channel becomes dominating.

The cases discussed so far correspond to superconducting elements having several valence orbitals per atom. A crucial test for the validity of our analysis would be provided by monovalent metals like Au and Na. They can be described by a single s orbital, and the model predicts a single channel with $T \approx 1$ for a one-atom contact^{16,17} in accordance with calculations based on molecular dynamics simulations²⁰. As these monovalent metals are not superconducting, our method of characterizing channels is not directly applicable. However, a small piece of normal metal in good contact with a large piece of a superconductor develops a gap in its quasiparticle density of states²¹. The lithographic mechanically controllable break junction technique³ allows for the fabrication of a constriction where the central region consists of Au atoms, embedded in a superconducting Al environment. Using shadow evaporation through a suspended mask we evaporated 400-nm-thick Al electrodes separated by a 50-nm-wide gap, which is filled with a 20-nm-thick Au layer deposited at a different angle under the same vacuum. The large asymmetry in thickness limits the influence of the normal metal on the quasiparticle density of states. As depicted in the lowermost panel of Fig. 1, we find indeed that the smallest Au contacts accommodate one single channel, contrary to the results for the sp -like and transition metals. However, we usually measure transmission values significantly below one. As for the other metals we interpret this reduced T with respect to the model predictions as arising from disorder in the contact. Our Au samples are presumably more prone to disorder because of the very small Au layer thickness. Considering that the accuracy in determining $\{T_n\}$ is also limited because of the modified quasiparticle density of states, the single channel fit is still very satisfactory in the case of Au.

We draw two conclusions from our results. First, they provide clear evidence that the smallest contacts produced by the different experimental techniques are indeed one-atom contacts, as N never exceeds N_{orb} . We stress that in these experiments a single atom determines a macroscopic quantity, that is, the total conductance of the macroscopic circuit in which it is embedded. The current which an one-atom contact can sustain can be as large as $100 \mu\text{A}$, corresponding to the very large current density of $10^{11} \text{ A cm}^{-2}$. This large current is carried by a small number of conduction channels. Second, the ensemble of our experimental results shows unambiguously that the conduction channels in an atomic contact are determined by the chemical nature of the central atom. As a crude rule of thumb, the number N of active channels corresponds to the number of valence electrons. This essential fact cannot be understood within a free electron model. Only a microscopic model that takes into account the atomic orbital structure as well as the local atomic geometry can fully explain the conduction properties of these contacts. We expect that these concepts will be applicable, in

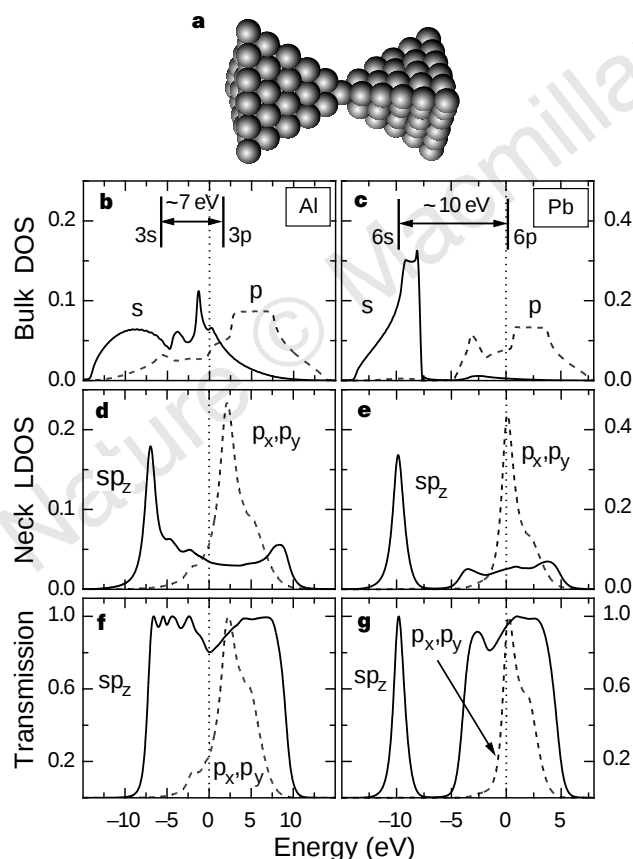


Figure 3 Localized orbitals model for electrical conduction through one-atom contacts. **a**, Model geometry. **b–g**, Schematic development of the atomic valence levels (thick vertical bars in **b** and **c**) into bulk conduction bands (**b** and **c**) for Al and Pb. The panels **d** and **e** depict schematically the local density of states (LDOS) in eV^{-1} at the central atom of the model geometry. The global energy dependence of the transmission coefficients T_n is shown in **f** and **g**. The dotted lines indicate the position of the Fermi level. The sp_z mode is the best transmitted mode for both materials. The p_x and p_y modes are degenerate due to the symmetry of the model geometry. The indicated energy differences between the atomic s and p levels are those used to fit the bulk band structure.

slightly modified version, to conduction through clusters of atoms and molecules. □

Received 29 December 1997; accepted 22 April 1998.

- Crommie, M. F., Lutz, C. P. & Eigler, D. M. Confinement of electrons to quantum corrals on a metal surface. *Science* **262**, 218–220 (1993).
- van Ruitenbeek, J. M. in *Mesoscopic Electron Transport* (eds Sohn, L. L., Kouwenhoven, L. P. & Schön, G.) 549–579 (Kluwer Academic, Dordrecht, 1997).
- van Ruitenbeek, J. M. *et al.* Adjustable nanofabricated atomic size contacts. *Rev. Sci. Instrum.* **67**, 108–111 (1996).
- Landauer, R. Electrical resistance of disordered one-dimensional lattices. *Phil. Mag.* **21**, 863–867 (1970).
- Scheer, E., Joyez, P., Esteve, D., Urbina, C. & Devoret, M. H. Conduction channel transmissions of atomic-size aluminum contacts. *Phys. Rev. Lett.* **78**, 3535–3538 (1997).
- Octavio, M., Tinkham, M., Blonder, G. E. & Klapwijk, T. M. Subharmonic energy-gap structure in superconducting constrictions. *Phys. Rev. B* **27**, 6739–6746 (1983).
- Averin, D. & Bardas, A. AC Josephson effect in a single quantum channel. *Phys. Rev. Lett.* **75**, 1831–1834 (1995).
- Cuevas, J. C., Martín-Rodero, A. & Levy Yeyati, A. Hamiltonian approach to the transport properties of superconducting quantum point contacts. *Phys. Rev. B* **54**, 7366–7379 (1996).
- Bratus, E. N., Shumeiko, V. S., Bezuglyi, E. V. & Wendin, G. dc-current transport and ac Josephson effect in quantum junctions at low voltage. *Phys. Rev. B* **55**, 12666–12677 (1997).
- Agrait, N., Rodrigo, J. G. & Vieira, S. Conductance steps and quantization in atomic-size contacts. *Phys. Rev. B* **47**, 12345–12348 (1996).
- Rubio, G., Agrait, N. & Vieira, S. Atomic-sized metallic contacts: mechanical properties and electronic transport. *Phys. Rev. Lett.* **76**, 2302–2305 (1996).
- Krans, J. M., van Ruitenbeek, J. M., Fisman, V. V., Yanson, I. K. & de Jongh, L. J. The signature of conductance quantization in metallic point contacts. *Nature* **375**, 767–769 (1995).
- Landman, U., Luedtke, W. D., Salisbury, B. E. & Whetten, R. L. Reversible manipulations of room temperature mechanical and quantum transport properties in nanowire junctions. *Phys. Rev. Lett.* **77**, 1362–1365 (1996).
- Todorov, T. N. & Sutton, A. P. Jumps in electronic conductance due to mechanical instabilities. *Phys. Rev. Lett.* **70**, 2138–2141 (1993).
- Beenakker, C. W. J. Quantum transport in semiconductor superconductor microjunctions. *Phys. Rev. B* **46**, 12841–12844 (1992).
- Cuevas, J. C., Levy Yeyati, A. & Martín-Rodero, A. Microscopic origin of the conducting channels in metallic atomic-size contacts. *Phys. Rev. Lett.* **80**, 1066–1069 (1998).
- Levy Yeyati, A., Martín-Rodero, A. & Flores, F. Conductance quantization and electron resonances in sharp tips and atomic size contacts. *Phys. Rev. B* **56**, 10369–10372 (1997).
- Lang, N. D. Resistance of atomic wires. *Phys. Rev. B* **52**, 5335–5342 (1995).
- Wan, C. C., Mozes, J.-L., Taraschi, G., Wang, J. & Guo, H. Quantum transport through atomic wires. *Appl. Phys. Lett.* **71**, 419–421 (1997).
- Brandbyge, M., Sørensen, M. R. & Jacobsen, K. W. Conductance eigenchannels in nanocontacts. *Phys. Rev. B* **56**, 14956–14959 (1997).
- Belzig, W., Bruder, C. & Schön, G. Local density of states in a dirty normal metal connected to a superconductor. *Phys. Rev. B* **54**, 9443–9448 (1996).

Acknowledgements. We thank D. Esteve and M. H. Devoret for discussions and C. Sürger for graphics preparation; N.A. and G.R.B. thank S. Vieira for discussions and support. The work was supported in part by the Deutsche Forschungsgemeinschaft (DFG), the Spanish CICYT, “Stichting FOM” (NWO) and the Bureau National de Métrologie (BNM).

Correspondence and requests for materials should be addressed to E.S. (e-mail: Elke.Scheer@phys.uni-karlsruhe.de).

Cation effects in doped La_2CuO_4 superconductors

J. P. Attfield, A. L. Kharlanov* & J. A. McAllister*

Interdisciplinary Research Centre in Superconductivity, University of Cambridge, Madingley Road, Cambridge CB3 0HE, UK, and Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, UK

* These authors contributed equally to the work

The critical temperatures of $(\text{Ln}_{1-x}\text{M}_x)_2\text{CuO}_4$ superconductors¹, in which Ln^{3+} (La and other lanthanides) and M^{2+} (Ca, Sr, Ba) cations are randomly distributed amongst the ‘type A’ lattice sites, are known to depend on the doping level, x , and the mean A-site cation radius, $\langle r_A \rangle$ (refs 2, 3). Here we show, by studying series of compositions with the same doping level and $\langle r_A \rangle$, that the critical temperature decreases linearly with increasing A-site disorder, as quantified by the variance in the distribution of A-site cation radii. From this, we are able to show that, in the absence of disorder, the critical temperature should increase quadratically with $\langle r_A \rangle$ for superconductors containing a single CuO_2 layer. Our results therefore show that the critical temperature is very sensitive to lattice strains, as has also been shown for the metal to insulator transition temperature in the magnetoresistive $(\text{Ln}_{1-x}\text{M}_x)\text{MnO}_3$ perovskites⁴.

Study of a series of $(\text{Ln}_{0.7}\text{M}_{0.3})\text{MnO}_3$ perovskites in which the mean A-site cation radius $\langle r_A \rangle$ was held constant has showed⁴ that the metal–insulator transition temperature T_m decreases linearly with the statistical variance σ^2 in the distribution of A-site radii⁵, $\sigma^2 = \langle r_A^2 \rangle - \langle r_A \rangle^2$. The hole-doped $(\text{Ln}_{1-x}\text{M}_x)_2\text{CuO}_4$ superconductors are similar to the above perovskites in that the doping level and lattice effects are controlled by the Ln^{3+} and M^{2+} cations at the A-sites, which in this case are adjacent to two-dimensional CuO_2 perovskite layers. Many compositions with the ‘optimum’ doping level $x = 0.075$, which maximizes the superconducting critical temperature T_c in the $(\text{La}_{1-x}\text{Sr}_x)_2\text{CuO}_4$ system⁶, have been reported, and the variation of T_c with $\langle r_A \rangle$ for series of $(\text{La}_{0.925}\text{Sr}_{0.075-x}\text{Ba}_x)_2\text{CuO}_4$ (ref. 7) and $(\text{La}_{0.925-x}\text{Ln}_x\text{Sr}_{0.075})_2\text{CuO}_4$ ($\text{Ln} = \text{Pr}, \text{Nd}, \text{Sm}, \text{Eu}$)^{3,8} compositions are shown in Fig. 1. Systematic changes with a maximum T_c at $\langle r_A \rangle = 1.22 \text{ \AA}$ are observed, but T_c is not uniquely determined by $\langle r_A \rangle$ at the given doping level.

To determine whether T_c is sensitive to A-site disorder, we have prepared three series of $(\text{La}_{0.925-x}\text{Nd}_x\text{Sr}_{0.075-g-h}\text{Ca}_g\text{Ba}_h)_2\text{CuO}_4$ compositions with the fixed $x = 0.075$ doping level (Table 1). Values of $\langle r_A \rangle$ and σ^2 were calculated from reported ionic radii for nine-coordination⁵. Initially, 15 compositions were prepared with $\langle r_A \rangle = 1.223 \text{ \AA}$, the value for $(\text{La}_{0.925}\text{Sr}_{0.075})_2\text{CuO}_4$ (the $f = g = h = 0$ composition). The A-cation size disorder varies between

Table 1 Properties of three A_2CuO_4 series with fixed mean A-site radius

A-site composition	σ^2 ($\text{\AA}^2$)	T_c (K)
$\langle r_A \rangle = 1.212 \text{ \AA}$ series		
$\text{La}_{0.899}\text{Nd}_{0.026}\text{Ca}_{0.075}$	0.0000	31.6
$\text{La}_{0.868}\text{Nd}_{0.057}\text{Ca}_{0.069}\text{Ba}_{0.006}$	0.0006	29.4
$\text{La}_{0.825}\text{Nd}_{0.100}\text{Ca}_{0.061}\text{Ba}_{0.014}$	0.0012	27.1
$\text{La}_{0.781}\text{Nd}_{0.144}\text{Ca}_{0.053}\text{Ba}_{0.022}$	0.0019	23.0
$\langle r_A \rangle = 1.223 \text{ \AA}$ series		
$\text{La}_{0.925}\text{Sr}_{0.075}$	0.0006	32.5
$\text{La}_{0.925}\text{Sr}_{0.060}\text{Ca}_{0.006}\text{Ba}_{0.007}$	0.0009	30.0
$\text{La}_{0.925}\text{Sr}_{0.045}\text{Ca}_{0.017}\text{Ba}_{0.013}$	0.0012	28.5
$\text{La}_{0.925}\text{Sr}_{0.030}\text{Ca}_{0.026}\text{Ba}_{0.020}$	0.0015	27.0
$\text{La}_{0.925}\text{Sr}_{0.015}\text{Ca}_{0.033}\text{Ba}_{0.027}$	0.0019	24.0
$\text{La}_{0.925}\text{Sr}_{0.008}\text{Ca}_{0.037}\text{Ba}_{0.030}$	0.0020	23.5
$\text{La}_{0.925}\text{Ca}_{0.041}\text{Ba}_{0.034}$	0.0022	20.0
$\text{La}_{0.900}\text{Nd}_{0.025}\text{Ca}_{0.037}\text{Ba}_{0.038}$	0.0025	20.5
$\text{La}_{0.875}\text{Nd}_{0.050}\text{Ca}_{0.032}\text{Ba}_{0.043}$	0.0029	18.0
$\text{La}_{0.850}\text{Nd}_{0.075}\text{Ca}_{0.028}\text{Ba}_{0.047}$	0.0033	16.0
$\text{La}_{0.825}\text{Nd}_{0.100}\text{Ca}_{0.023}\text{Ba}_{0.052}$	0.0036	–
$\text{La}_{0.800}\text{Nd}_{0.125}\text{Ca}_{0.019}\text{Ba}_{0.056}$	0.0040	–
$\text{La}_{0.775}\text{Nd}_{0.150}\text{Ca}_{0.014}\text{Ba}_{0.061}$	0.0043	–
$\text{La}_{0.750}\text{Nd}_{0.175}\text{Ca}_{0.009}\text{Ba}_{0.066}$	0.0047	–
$\text{La}_{0.725}\text{Nd}_{0.200}\text{Ca}_{0.005}\text{Ba}_{0.070}$	0.0050	–
$\langle r_A \rangle = 1.232 \text{ \AA}$ series		
$\text{La}_{0.925}\text{Sr}_{0.019}\text{Ba}_{0.056}$	0.0035	32.0
$\text{La}_{0.921}\text{Nd}_{0.004}\text{Sr}_{0.018}\text{Ba}_{0.057}$	0.0036	31.7
$\text{La}_{0.906}\text{Nd}_{0.019}\text{Sr}_{0.013}\text{Ba}_{0.062}$	0.0039	31.1
$\text{La}_{0.888}\text{Nd}_{0.037}\text{Sr}_{0.007}\text{Ba}_{0.068}$	0.0043	30.0
$\text{La}_{0.868}\text{Nd}_{0.058}\text{Ba}_{0.075}$	0.0047	29.0

Here σ^2 is the cation size variance, and $\langle r_A \rangle$ is the fixed mean A-site radius. A dash in the rightmost column indicates that superconductivity was not detected.

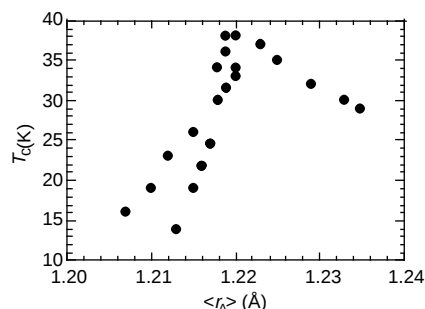


Figure 1 Plot of T_c for previously reported $(\text{Ln}_{0.925}\text{M}_{0.075})_2\text{CuO}_4$ materials^{3,7,8} against average A-cation radius $\langle r_A \rangle$.

$\sigma^2 = 0.0006 \text{ \AA}^2$ for the latter sample and $0.0050 \text{ \AA}^2$ for $(\text{La}_{0.725}\text{Nd}_{0.200}\text{Ca}_{0.005}\text{Ba}_{0.070})_2\text{CuO}_4$. Ceramic samples were sintered under identical conditions in air at $1,050^\circ\text{C}$ for 200 h, quenched to room temperature to avoid A-cation segregation, then annealed in flowing oxygen gas at 400°C for 50 h. Room-temperature powder X-ray diffraction showed all the samples to be single phase with $I4/mmm$ tetragonal symmetry for $\sigma^2 < 0.0038 \text{ \AA}^2$ and the orthorhombic $Abma \sqrt{2a} \times \sqrt{2a} \times c$ superstructure above this limit. The orthorhombic-tetragonal transition temperature T_{ot} is 180 K for $(\text{La}_{0.925}\text{Sr}_{0.075})_2\text{CuO}_4$ (ref. 9) so that between 180 and 300 K the average $dT_{ot}/d\sigma^2$ is $\approx 38,000 \text{ K \AA}^{-2}$. Alternating current susceptometry measurements on powdered samples in an applied field of 1 G showed a systematic decrease in T_c (Fig. 1) and in the diamagnetic volume fraction with increasing σ^2 , from 20% of perfect diamagnetism at $\sigma^2 = 0.0006 \text{ \AA}^2$ to zero at $\sigma^2 = 0.0035 \text{ \AA}^2$. Thermogravimetric analysis (TGA) showed the oxygen contents of all the samples to be 4.01 ± 0.02 and no systematic changes were observed across the series. However, the T_c of 32.5 K for the first member of the series, $(\text{La}_{0.925}\text{Sr}_{0.075})_2\text{CuO}_4$, is slightly lower than the values of 36–38 K reported elsewhere^{7,9} suggesting a slight oxygen deficiency in these samples, below the sensitivity of the TGA. Figure 2 shows that T_c decreases linearly with σ^2 , with an average deviation of $<1 \text{ K}$ for the 11 values. Fitting this as:

$$T_c = T_c^0 - p\sigma^2 \quad (1)$$

gives the extrapolated critical temperature at zero variance $T_c^0 = 35.8(2) \text{ K}$ and the negative slope as $p = 6,100(110) \text{ K \AA}^{-2}$, with fitting errors in parenthesis.

The series of compositions with larger mean radius $\langle r_A \rangle = 1.232 \text{ \AA}$, prepared as above, is limited to the disorder range to $0.0035 < \sigma^2 < 0.0047 \text{ \AA}^2$ by the radii of the available cations. T_c again shows a linear dependence upon σ^2 (Fig. 2) with fitted parameters $T_c^0 = 40.6(4) \text{ K}$ and $p = 2,450(90) \text{ K \AA}^{-2}$.

The preparation of $(\text{Ln}_{0.925}\text{M}_{0.075})_2\text{CuO}_4$ samples with a small mean radius, $\langle r_A \rangle = 1.212 \text{ \AA}$, has proved more difficult as they require higher Ca contents. At atmospheric pressure this leads to the formation of a secondary phase, $\text{La}_2\text{CaCu}_2\text{O}_6$, and so four samples (Table 1), initially prepared as before, were heated at 900°C under 550 atm O_2 pressure for 6 h to improve homogeneity^{10,11}. However, a trace of the secondary phase ($\sim 1\%$ by mass) is still observed in the most Ca-rich sample, and this residual inhomogeneity accounts for the systematic but nonlinear decrease of T_c with σ^2 (Fig. 2). T_c^0 is taken to be 31.6 K , the value for the first sample which has $\sigma^2 < 0.0001 \text{ \AA}^2$. We also note that the outlying T_c value in the $\langle r_A \rangle = 1.223 \text{ \AA}$ series (Fig. 2) is for the most Ca-rich composition, $(\text{La}_{0.925}\text{Ca}_{0.041}\text{Ba}_{0.034})_2\text{CuO}_4$. This illustrates the need for very cation-homogeneous samples to establish σ^2 trends.

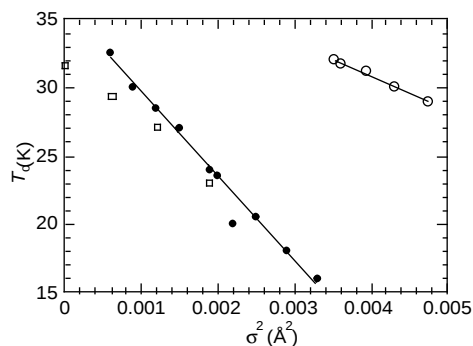


Figure 2 Variation of T_c with A-cation size variance σ^2 for three series of $(\text{Ln}_{0.925}\text{M}_{0.075})_2\text{CuO}_4$ samples. Each series has a fixed $\langle r_A \rangle$ as follows; $1.212 \text{ \AA}$ (open squares), $1.223 \text{ \AA}$ (filled circles) and $1.232 \text{ \AA}$ (open circles). Linear fits to the last two series are shown.

The above results show that at a constant doping level and A-cation size $\langle r_A \rangle$, T_c decreases systematically with the A-cation size variance σ^2 . A linear dependence is observed in the two series of homogeneous samples, and T_c increases and p decreases with increasing $\langle r_A \rangle$. These size and disorder effects account for the variation of previously observed T_c s with $\langle r_A \rangle$ in Fig. 1. Below $1.22 \text{ \AA}$, the A-sites contain mixtures of Ln^{3+} and Sr^{2+} cations which are similar in size so that the disorder contribution is small and T_c increases strongly with $\langle r_A \rangle$ up to a maximum close to the radius of La ($1.216 \text{ \AA}$). Increasing $\langle r_A \rangle$ above this value requires substitution of the large Ba^{2+} cation ($r = 1.47 \text{ \AA}$) for Sr^{2+} , giving rise to an increasing A-cation size variance σ^2 which offsets the increase in average size and causes T_c to decline. $(\text{La}_{0.925}\text{Ba}_{0.075})_2\text{CuO}_4$ has the largest $\langle r_A \rangle$ and σ^2 , and T_c is reduced by $\sim 10 \text{ K}$ due to the size mismatch between the cations. Larger $\langle r_A \rangle$ values are observable in the oxyfluorides $(\text{Sr,Ba})_2\text{CuO}_2\text{F}_{2+\delta}$ which are isostructural with the A_2CuO_4 materials, but require only M^{2+} cations at the A sites, resulting in higher T_c s up to 64 K (ref. 12).

The A_2CuO_4 materials offer the ideal opportunity to study the effects of cation disorder in single copper oxide layer superconductors, as the A-sites can accommodate many cations and other types of disorder are minimal. However, they can only be prepared over a small range of $\langle r_A \rangle$. To estimate T_c for optimally doped superconductors at larger $\langle r_A \rangle$ in the absence of lattice disorder is difficult for other structures, but in the special case when the A-sites are occupied by only one type of cation (Sr or Ba), then the maximum observed T_c from the known single CuO_2 layer materials may be taken as T_c^0 for that A-cation radius. The highest reported T_c for a single-layer material containing only Sr^{2+} ($r_A = 1.31 \text{ \AA}$) at the A sites is 70 K for $\text{Tl}_{0.5}\text{Pb}_{0.5}\text{Sr}_4\text{Cu}_2(\text{CO}_3)\text{O}_7$ (ref. 13). $\text{Tl}_2\text{Ba}_2\text{CuO}_{6+\delta}$ ($T_c = 93 \text{ K}$; ref. 14) and $\text{HgBa}_2\text{CuO}_{4+\delta}$ ($T_c = 97 \text{ K}$; ref. 15) both have only Ba^{2+} ($r = 1.47 \text{ \AA}$) cations at the A-sites and the latter provides an estimate of T_c^0 at this radius. Plotting T_c^0 values with those from the three $(\text{Ln}_{0.925}\text{M}_{0.075})_2\text{CuO}_4$ series shows that the critical temperatures of near-optimally doped, disorder-free, single-layer copper oxide superconductors increase smoothly with $\langle r_A \rangle$ (Fig. 3). This variation can be described by a quadratic form analogous to equation (1);

$$T_c^0 = T_c^m - q(r_A^m - \langle r_A \rangle)^2 \quad (2)$$

where the maximum T_c^m is found for maximum A-site radius r_A^m . The fitted parameters are $T_c^m = 98(1) \text{ K}$, $r_A^m = 1.49(1) \text{ \AA}$, $q = 870(30) \text{ K \AA}^{-2}$, although these values are rather approximate given the uncertainties in the T_c^0 data. Hence, it is possible that higher T_c superconductors may be obtained if structures containing larger A-site cations such as K^+ ($r = 1.55 \text{ \AA}$) can be stabilized.

The T_c variations of optimally doped $(\text{Ln}_{0.925}\text{M}_{0.075})_2\text{CuO}_4$ materials and other single CuO_2 layer superconductors at ambient

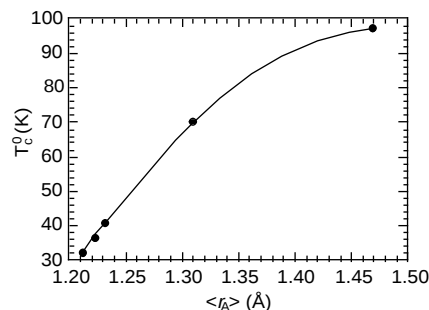


Figure 3 Plot of the critical temperature corrected for disorder, T_c^0 , against $\langle r_A \rangle$ for single CuO_2 layer superconductors, with a fit of equation (2).

pressure are consistent with a maximum $T_c \approx 100$ K, decreased by reductions in the A-cation size and disorder at the A (and perhaps other) sites. The quadratic dependence of T_c upon radius in equations (1) and (2) shows that lattice strains are responsible for these variations. σ in equation (1) and $(r_A^m - \langle r_A \rangle)$ in equation (2) are complementary measures of the displacements of adjacent oxygen atoms due to A-cation disorder and size which result in deformations and rotations of the linked CuO_6 octahedra.

The changes of transition temperatures with σ^2 reflect their sensitivity to lattice effects and give a measure of the accompanying lattice distortion. In our $(\text{Ln}_{0.925}\text{M}_{0.075})_2\text{CuO}_4$ series with $\langle r_A \rangle = 1.223$ Å, the purely structural orthorhombic–tetragonal change, which is a typical soft mode transition at which a static lattice distortion occurs¹⁶, has an average $dT_o/d\sigma^2 \approx 38,000 \text{ K Å}^{-2}$. $dT_m/d\sigma^2 \approx -21,000 \text{ K Å}^{-2}$ for the metal–insulator transition in the $(\text{Ln}_{0.7}\text{M}_{0.3})\text{MnO}_3$ manganites is smaller but comparable in magnitude to the former value, reflecting the formation of large static Jahn–Teller distortions that accompanies the localization of carriers at T_m (ref. 4,17). The superconducting transition in the copper oxides has still smaller values of $dT_c/d\sigma^2$ from $-6,100$ to $-2,500 \text{ K Å}^{-2}$, consistent with a small structural distortion at T_c , which has been observed in some materials, for example, $\text{HgBa}_2\text{CuO}_{4+\delta}$ (ref. 18).

We have shown that the use of atomic size variance, in addition to the mean atomic size, allows the effects of the A-type cations upon T_c to be investigated systematically at a fixed doping level in A_2CuO_4 superconductors. T_c is found to vary linearly with lattice strains. This approach should enable the lattice and electronic contributions to be separated for series in which the doping also changes, such as $(\text{La}_{1-x}\text{Sr}_x)_2\text{CuO}_4$ with variable x . Representing physical properties as a function of x , $\langle r_A \rangle$ and σ^2 should also allow data from all $(\text{Ln}_{1-x}\text{M}_x)_2\text{CuO}_4$ families to be combined into a single phase diagram. These methods may be applicable to other electronic transitions in disordered crystalline materials. □

Received 20 February; accepted 8 April 1998.

1. Bednorz, J. G. & Müller, K. A. Possible high- T_c superconductivity in the Ba-La-Cu-O system. *Z. Phys. B* **64**, 189–193 (1986).
2. Torrance, J. B. *et al.* Properties that change as superconductivity disappears at high-doping concentrations in $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$. *Phys. Rev. B* **40**, 8872–8877 (1989).
3. Tarascon, J. M., Greene, L. H., McKinnon, W. R. & Hull, G. W. Superconductivity in rare-earth doped oxygen-defect perovskites $\text{La}_{2-x}\text{Sr}_x\text{CuO}_{4-y}$. *Solid State Commun.* **63**, 499–505 (1987).
4. Rodríguez-Martínez, L. M. & Attfield, J. P. Cation disorder and size effects in magnetoresistive manganese oxide perovskites. *Phys. Rev. B* **54**, R15622–R15625 (1996).
5. Shannon, R. D. Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Crystallogr. A* **32**, 751–767 (1976).
6. Tarascon, J. M., Green, L. H., McKinnon, W. R., Hull, G. W. & Geballe, T. H. Superconductivity at 40 K in the oxygen-defect perovskites $\text{La}_{2-x}\text{Sr}_x\text{CuO}_{4-y}$. *Science* **235**, 1373–1376 (1987).
7. Oh-Ishii, K. & Syono, Y. Dependence of lattice parameters and T_c on the hole concentration determined by precise measurement of the oxygen content of $\text{M}_x\text{La}_{2-x}\text{CuO}_{4-y}$ ($\text{M} = \text{Ba}, \text{Sr}, \text{Ca}$). *J. Solid State Chem.* **95**, 136–144 (1991).
8. Büchner, B. *et al.* Low temperature phase transition and superconductivity in $(\text{La Nd})\text{-Sr-Cu-O}$. *Physica C* **185–189**, 903–904 (1991).
9. Day, P. *et al.* Temperature dependence of the crystal structure of the ceramic superconductor $\text{La}_{1.8}\text{Sr}_{0.15}\text{CuO}_4$: a powder neutron diffraction study. *J. Phys. C: Solid State Phys.* **20**, L429–L434 (1987).
10. Dabrowski, B. *et al.* Suppression of superconducting transition temperature in orthorhombic $\text{La}_{2-x}\text{Ca}_x\text{CuO}_4$. *Physica C* **217**, 455–460 (1993).
11. Dabrowski, B. *et al.* Dependence of superconducting transition-temperature on doping and structural distortion of the CuO_2 planes in $\text{La}_{2-x}\text{M}_x\text{CuO}_4$ ($\text{M} = \text{Nd}, \text{Ca}, \text{Sr}$). *Phys. Rev. Lett.* **76**, 1348–1351 (1996).
12. Slater, P. R. *et al.* An improved route to the synthesis of superconducting copper oxyfluorides $\text{Sr}_{2-x}\text{A}_x\text{CuO}_2\text{F}_{2+8}$ ($\text{A} = \text{Ca}, \text{Ba}$) using transition metal difluorides as fluorinating reagents. *Physica C* **253**, 16–22 (1995).
13. Huve, M. *et al.* A 70 K superconductor, the oxycarbonate $\text{Ti}_{0.5}\text{Pb}_{0.5}\text{Sr}_4\text{Cu}_2(\text{CO}_3)_7\text{O}_7$. *Physica C* **205**, 219–224 (1993).
14. Wagner, J. L. *et al.* Multiple defects in overdoped $\text{Tl}_2\text{Ba}_2\text{CuO}_{6+\delta}$: effects on structure and superconductivity. *Physica C* **277**, 170–182 (1997).
15. Asab, A., Armstrong, A. R., Gemeson, I. & Edwards, P. P. Single-step synthesis and crystal structure of $\text{HgBa}_2\text{CuO}_{4+\delta}$ with a T_c of 97 K. *Physica C* **255**, 180–187 (1995).
16. Axe, J. D. & Crawford, M. K. Structural instabilities in lanthanum cuprate superconductors. *J. Low Temp. Phys.* **95**, 271–284 (1994).
17. Rodríguez-Martínez, L. M. & Attfield, J. P. Cation disorder and the metal–insulator transition temperature in manganese oxide perovskites. *Phys. Rev. B* (in the press).
18. Bordet, P. *et al.* Structural instability around T_c observed in Hg-1201 by neutron powder diffraction and EXAFS. *Physica C* **282**, 1081–1082 (1997).

Acknowledgements. We thank H. Fuess for access to thermogravimetric equipment at Technische Universität Darmstadt, and the British Council for an ARC grant.

Correspondence and requests for materials should be addressed to J.P.A. (e-mail: jpa14@cam.ac.uk).

A molecular metal with ion-conducting channels

Takayoshi Nakamura*, Tomoyuki Akutagawa*, Kazumasa Honda†, Allan E. Underhill‡, A. Treeve Coomber§ & Richard H. Friend§

* Research Institute for Electronic Science, Hokkaido University, Sapporo 060, Japan

† National Institute for Materials and Chemical Research, Tsukuba, Ibaraki 305, Japan

‡ Department of Chemistry, University College of North Wales, Bangor, Gwynedd LL57 2UW, UK

§ Cavendish Laboratory, University of Cambridge, Madingley Road, Cambridge CB3 0HE, UK

Metallic behaviour is well known in charge-transfer complexes that contain stacks of planar, partially oxidized (or reduced) π -conjugated molecules. Electronic conduction occurs in the partially occupied, delocalized π bands formed by intermolecular orbital overlap, and some of these materials exhibit superconductivity^{1,2}. Counter-ions, present to achieve charge neutrality, usually play a passive role, although in some cases they couple to the electronic structure, for example by imposing a new structural periodicity (a superlattice) by orientational ordering¹. The development of molecular solids that can simultaneously support the transport of both electrons and ions is important for several fields, including the development of solid-state batteries^{3,4}, electroluminescent devices⁵ and biomimetic systems^{6,7}. Crown ethers are promising components for such systems, as they provide cavities through which ion motion might occur. Here we report that the charge-transfer salt $\text{Li}_{0.6}(\text{15-crown-5-ether})[\text{Ni}(\text{dmit})_2]_2 \cdot \text{H}_2\text{O}$ exhibits both electron and ion conductivity: the stacks of the nickel complex (dmit is an organic molecule) provide a pathway for electron conduction, and stacks of the crown ethers provide channels for lithium-ion motion. Evidence for the latter above 250 K is provided by NMR and conductivity studies. We also see evidence for coupling of the electron and ion motions. This compound might serve as a model for the development of other hybrid electronic/ionic conducting materials.

We have discovered that crown ethers can be readily incorporated in a range of charge-transfer salts formed with the planar, π -conjugated acceptor complex, $\text{Ni}(\text{dmit})_2$, the structure for which is shown in Fig. 1a (here dmit is 2-thioxo-1,3-dithion-4,5-dithiolate). This complex is well-known as an electron acceptor, and has been used to produce many metallic conductors, and even superconductors⁸. For the salt reported here, the crown ethers provide channels in which the lithium ions are contained and which run parallel to the $\text{Ni}(\text{dmit})_2$ stacks.

Crystals of composition $\text{Li}_{0.6}(\text{15-crown-5-ether})[\text{Ni}(\text{dmit})_2]_2 \cdot \text{H}_2\text{O}$ were grown by electrocrystallization of an acetonitrile solution of tetrabutylammonium- $\text{Ni}(\text{dmit})_2$ and LiClO_4 in the presence of 15-crown-5 in an H-cell using platinum electrodes at constant current. A mixture of needles and small platelets was obtained, and these were found to be crystallographically identical. Figure 1b shows the ion channels and the $\text{Ni}(\text{dmit})_2$ stacks which extend along the a axis, and which provide for transport of ions (Li^+) and electrons, respectively. Figure 1c illustrates that at room temperature the stack of $\text{Ni}(\text{dmit})_2$ is regular, so the partially-occupied π conduction band formed by intermolecular overlap of π orbitals along this stack is expected to support metallic behaviour. The intermolecular overlap is shown in Fig. 1d. The overlap shows a longitudinal offset, giving a metal-over-ring arrangement which is favourable for strong intermolecular interactions. The face-to-face

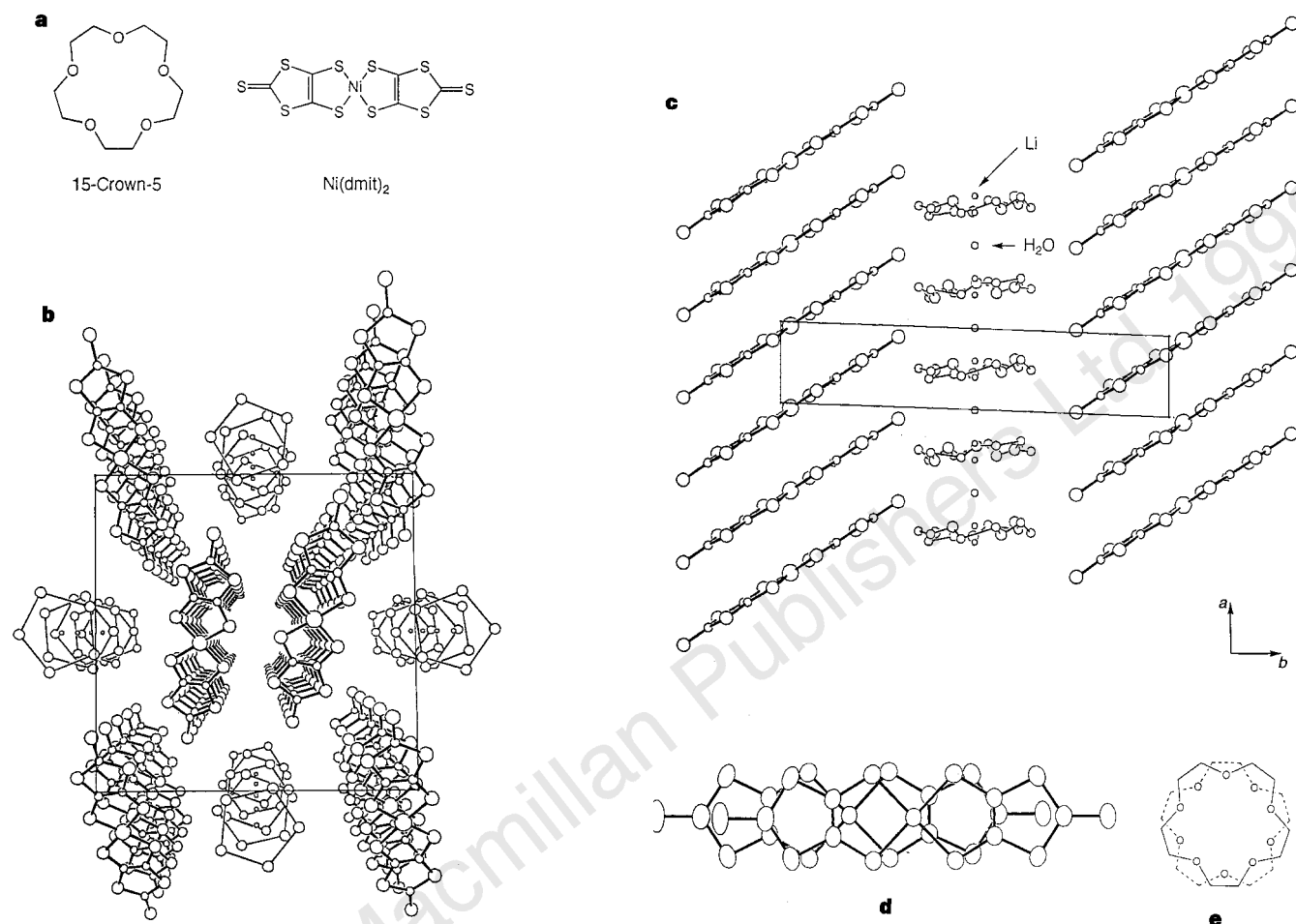


Figure 1 Structure and packing of the 15-crown-5-ether and Ni(dmit)_2 . **a**, Chemical structures. **b**, Projection of the crystal structure of $\text{Li}_{0.6}$ (15-crown-5-ether) $[\text{Ni(dmit)}_2]_2 \cdot \text{H}_2\text{O}$ along the a axis. As there is disorder in the crown ether, only the oxygen atom positions are shown. **c**, Crystal structure in the a - b plane. X-ray analysis indicates an unknown electron-density peak at the mid-point of the centres of gravity of neighbouring crown-ether molecules in the column. This is attributed to oxygen (present as water) so that the calculated values of the elemental composition and crystal density were in good agreement with those observed. Lithium cations are distributed just above and just below the centre of gravity of the crown-ether molecules. The Li:Ni ratio was determined to be 0.3 by

distance between adjacent Ni(dmit)_2 moieties within the column is 3.70 Å, which is somewhat longer than that observed in the superconductor $\text{Me}_4\text{N}^+ (\text{Ni(dmit)}_2)_2$ (ref. 9). Ni(dmit)_2 columns are separated from one another by the macrocyclic molecules which prevent the formation of short sulphur-sulphur contacts between ligands of adjacent pairs of Ni(dmit)_2 , thus resulting in a highly one-dimensional electronic structure for the crystal. The crown ether macrocycles form an almost regular stack, creating ion channels parallel to the electronically conducting molecular stacks as shown in Fig. 1e. The crown ethers are disordered, possessing two orientations related to the point symmetry around the centre of gravity of the molecule. One water molecule occupies a position at the mid-point of two adjacent crown ether molecules. The lithium cations are located in positions just above and just below the centre of the crown-ether plane, thus forming a mixed chain in the ion channel with the water molecule. From the measured chemical composition we note that occupancy of the lithium sites is 0.3.

The electronic properties of this material show unusual behaviour, with a crossover from a metallic state near room temperature—associated with delocalized electrons in the conduction band of the Ni(dmit)_2 stacks—to a magnetic insulator state below

inductively coupled plasma mass spectroscopy, so that the occupancy of each Li site is 0.3. The crystal data are summarized as follows: monoclinic space group $P2_1/c$, $a = 4.408 \text{ Å}$, $b = 21.65 \text{ Å}$ and $c = 21.38 \text{ Å}$, $V = 2039 \text{ Å}^3$; $D_c = 1.86$ and $D_m = 1.86 \text{ g cm}^{-3}$ for $z = 2$ with $\text{Li}_{0.5}(\text{C}_{10}\text{H}_{20}\text{O}_5) \cdot (\text{C}_6\text{S}_{10}\text{Ni})_2 \cdot \text{H}_2\text{O}$. The data collection was performed on a Rigaku AFC-5 4-circle diffractometer with graphite-monochromated $\text{Mo K}\alpha$ radiation. The structure was solved by a direct method¹⁷. Further crystal data are available: see Supplementary Information. **d**, The overlap between adjacent Ni(dmit)_2 molecules in the column. **e**, The stacking of the crown ethers along the column. These are present in two orientations.

~200 K. We attribute this crossover to the interaction between the Li^+ ions in the crown-ether columns and the electrons present in the Ni(dmit)_2 stacks.

The salt at room temperature shows a high, metallic level of electrical conductivity, which is measured to be 240 S cm^{-1} along the a axis. This is as expected for a uniform stack structure. On cooling, the conductivity rises slowly, reaching a maximum near 250 K, before falling at lower temperatures, as is shown in Fig. 2. The low-temperature behaviour can be fitted with an activated law, with activation energy 12 meV. Measurements were also made under hydrostatic pressure, using techniques described elsewhere^{10,11}. The room-temperature conductivity was found to increase with increasing pressure, at a rate of $9.1\% \text{ kbar}^{-1}$, but the low-temperature fall in conductivity was not removed, although the activation energy was reduced to 2.4 meV at 10 kbar.

The clue to the cause of the non-metallic low-temperature behaviour is given by the temperature-dependent magnetic susceptibility, shown in Fig. 3, which provides evidence for antiferromagnetically coupled localized spins at low temperatures. The low-temperature behaviour (below 200 K) is quantitatively modelled as an antiferromagnetically coupled Heisenberg chain¹², with an

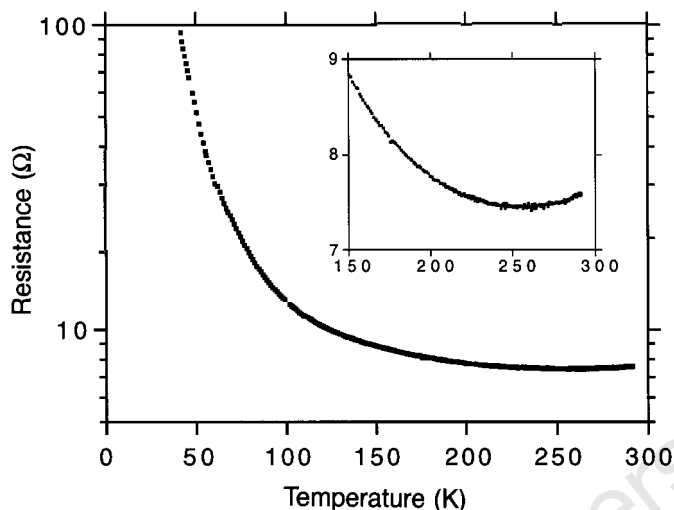


Figure 2 Temperature dependence of electrical resistance of a crystal of $\text{Li}_{0.6}$ (15-crown-5-ether) $[\text{Ni}(\text{dmit})_2]_2 \cdot \text{H}_2\text{O}$. Needle crystals (typical dimensions $0.1 \times 0.05 \times 4$ mm) were measured, with four in-line evaporated gold contacts

and a four-terminal conductivity measurement. The inset shows the conductivity maximum near 250 K.

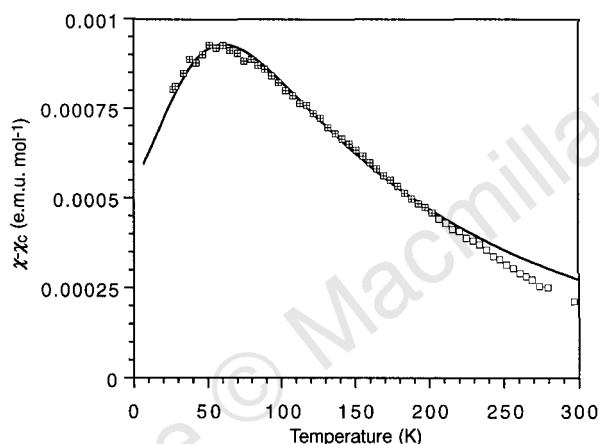


Figure 3 Magnetic susceptibility versus temperature for $\text{Li}_{0.6}$ (15-crown-5-ether) $[\text{Ni}(\text{dmit})_2]_2 \cdot \text{H}_2\text{O}$. The measured data (χ) have been corrected for a small paramagnetic tail (χ_0), with a Curie constant equivalent to 1% of $g = 2$, spin-1/2 species. The solid line is a fit of the data below 200 K (squares with cross inside) to an expression describing a Heisenberg antiferromagnetic chain, with an exchange energy, J/k_B of 90 K and 0.45 spins per crown-ether unit¹². Measurements were performed on polycrystalline samples, using a Faraday balance.

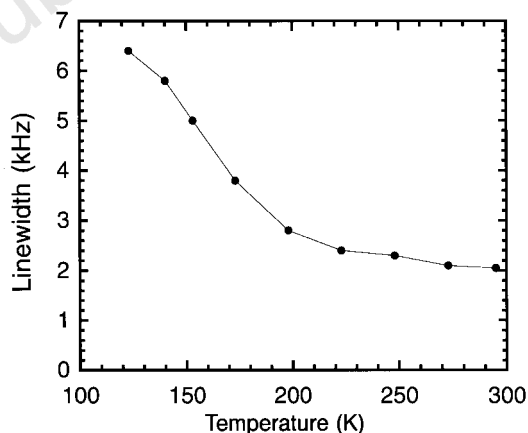


Figure 4 Temperature dependence of the linewidth of cross-polarization magic-angle-spinning NMR spectra for ^7Li in crystals of $\text{Li}_{0.6}$ (15-crown-5-ether) $[\text{Ni}(\text{dmit})_2]_2 \cdot \text{H}_2\text{O}$. The linewidth is given as full-width at half-maximum: the spectra were measured with a Hemagnetics CNX400-NMR system.

exchange energy, J/k_B , of 90 K and 0.45 spins ($g = 2$) per crown-ether unit. The exchange energy is related to the intermolecular transfer integral, t , and the on-site Coulomb repulsion energy, U , by $J = 2t^2/U$; values of exchange energy of this size are commonly found in charge-transfer salts of this type, for which values of U of ~ 1 eV and t of ~ 0.1 eV are expected¹³. This magnetic behaviour is consistent with localization of the electrons on the $\text{Ni}(\text{dmit})_2$ stacks which balance the charge on the Li^+ ions. Above 200 K the magnetic susceptibility falls more quickly than modelled, as is clearly seen in Fig. 3. We consider that this is due to a gradual transition from a localized Heisenberg antiferromagnetic state to the metallic state which is present above 250 K, and for which a Pauli susceptibility is expected.

We expect that the Li^+ ions present in the crown-ether columns should be mobile at elevated temperatures, given that the occupancy of each Li^+ ion site is only 0.3. We have made measurements of the solid-state NMR relaxation spectra for ^7Li in crystals, and find that

there is substantial narrowing of the line at temperatures above 200 K, as is shown in Fig. 4. This narrowing of the line with increasing temperature suggests that the lithium ions in the crystal are diffusing at temperatures above 200 K, producing motional narrowing. Rotation of the crown ether might also contribute, but we consider that the large difference in the quadrupolar coupling between 123 K (~ 17 kHz) and room temperature (~ 8 kHz) provides strong support for the model of lithium-ion motion. The quadrupole coupling depends on the electric field gradient¹⁴. At the inversion centres of the crystal—for example, the centre of gravity of the crown ether—the electric field gradient should be zero. The field gradient should be relatively large at the Li^+ positions just above and just below the 15-crown-5 molecules indicated in Fig. 1c, resulting in the large quadrupole coupling. The smaller quadrupole coupling at higher temperatures indicates that the Li^+ ions are passing (within the NMR time constant) through the centres of the crown ether with zero field gradient, as

well as through the large-field-gradient positions just above and below the crown-ether molecules.

We have also made direct measurements of the ionic conductivity of the crown-ether salt and also of reference materials. We used an electron-blocking method, in which a compacted disk of the material under test was sandwiched between Li^+ /polyethylene oxide disks which were pressed onto Li metal electrodes. The ionic conductivity of $(\text{PEG})\text{-Li}(\text{CF}_3\text{SO}_3)$, where PEG is polyethylene glycol, was measured to be $1.6 \times 10^{-6} \text{ S cm}^{-1}$ at 60°C , in good agreement with literature values¹⁵. We measured values, also at 60°C , for a range of $\text{Ni}(\text{dmit})_2$ salts: $3 \times 10^{-6} \text{ S cm}^{-1}$ for $\text{Li}_{0.6}$ (15-crown-5-ether) $[\text{Ni}(\text{dmit})_2]_2\text{H}_2\text{O}$; $2 \times 10^{-8} \text{ S cm}^{-1}$ for $(\text{C}_{16}\text{H}_{33})_2(\text{CH}_3)_2\text{N}[\text{Ni}(\text{dmit})_2]$; and $3 \times 10^{-7} \text{ S cm}^{-1}$ for $(\text{TPP})[\text{Ni}(\text{dmit})_2]_3$, where TPP is tetraphenylphosphonium. Each measurement was made after the current reached its saturated value, typically between 30 and 60 minutes after application of the constant voltage. We have checked that there is no decomposition (for example, water loss) at 60°C (by thermal gravimetry), and consider that the value of ionic conductivity for the crown-ether Li^+ salt indicates clearly that there is significant Li^+ mobility in this material at this temperature. We have measured significantly lower values in the (electronically insulating) tetra-alkyl ammonium salt and the (electronically conducting) tetraphenylphosphonium salt.

We consider that the electron dynamics on the $\text{Ni}(\text{dmit})_2$ chains are controlled by the dynamics of the Li^+ ions in the crown-ether columns. Above 250 K the Li^+ ions are mobile, and do not localize the conduction electrons on the $\text{Ni}(\text{dmit})_2$ stacks, which therefore show metallic properties. Between 250 and 200 K the Li^+ ions become immobilized, and provide static coulombic pinning potentials for the conduction electrons on the $\text{Ni}(\text{dmit})_2$ stacks, causing localization and the magnetic insulator properties evident in Figs 2 and 3. A metal-insulator transition driven by interaction between ions and electrons in this way is unusual, and the example that we have illustrated here may provide insights into transport properties of other mixed ionic/electronic conductors, such as the conjugated polymer blends recently used by Pei *et al.*⁵ in electroluminescent diodes.

The electron-transport system with ion channels that we describe here may allow for the design of electron-cation transport systems for use in active and selective ion transport. For example, a double carrier process could be realised in a redox gradient where the coupled parallel transport of electrons and alkali-metal cations takes place¹⁶. □

Received 17 April; accepted 13 May 1998.

- Jérome, D. The physics of organic superconductors. *Science* **252**, 1509–1515 (1991).
- Williams, J. M. *et al.* Organic superconductors—new benchmarks. *Science* **252**, 1501–1508 (1991).
- Takahashi, T. *et al.* Solid-state ionics—recent trends and future expectations. *Bull. Electrochem.* **11**, 1–33 (1995).
- Fauteux, D., Massucco, A., McIn, M., Vanburen, M. & Shi, J. Lithium polymer electrolyte rechargeable battery. *Electrochim. Acta* **40**, 2185–2190 (1995).
- Pei, Q., Yu, G., Zhang, C., Yang, Y. & Heeger, A. J. Polymer light-emitting electrochemical cells. *Science* **269**, 1086–1089 (1995).
- Pregel, M. J., Jullien, L. & Lehn, J. M. Towards artificial ion channels—transport of alkali-metal ions across liposomal membranes by bouquet molecules. *Angew. Chem. Int. Edn Engl.* **31**, 1637–1640 (1992).
- Ghadiri, M. R., Granja, J. R. & Buehler, L. K. Artificial transmembrane ion channels from self-assembling peptide nanotubes. *Nature* **369**, 301–304 (1994).
- Cassoux, P. *et al.* Molecular-metals and superconductors derived from metal-complexes of 1,3-dithiol-2-thione-4,5-dithiolate (dmit). *Coord. Chem. Rev.* **110**, 115–160 (1991).
- Kim, H., Kobayashi, A., Sasaki, Y., Kato, R. & Kobayashi, H. New radical-anion complex, $[(\text{CH}_3)_4\text{N}][\text{Ni}(\text{dmit})_2]$ with metal-semimetal phase-transition. *Chem. Lett.* 1799–1802 (1987).
- Guy, D. R. P. & Friend, R. H. Temperature measurement in high pressure cells using a rhodium + 0.5% iron versus chromel thermocouple pair. *J. Phys. E* **19**, 430–433 (1986).
- Murata, K. *et al.* Superconductivity with the onset at 8 K in the organic conductor beta-(BEDT-TTF)₂ under pressure. *J. Phys. Soc. Jpn* **54**, 1236–1239 (1985).
- Estes, W. E., Gavel, D. P., Hatfield, W. E. & Hodgson, D. J. Magnetic and structural characterisation of dibromo- and dichlorobis(thiazole) copper(II). *Inorg. Chem.* **17**, 1415–1421 (1978).
- Obertelli, S. D., Friend, R. H., Talham, D. R., Kurmoo, M. & Day, P. Magnetic susceptibility and EPR of the organic conductors α' -(BEDT-TTF)₂X, X=AuBr₂, CuCl₂ and Ag(CN)₂. *J. Phys. Cond. Matter* **1**, 5671–5680 (1989).
- Cohen, M. H. & Reif, F. in *Solid State Physics* (eds Seitz, F. & Turnbull, D.) 321–438 (Academic, New York, 1957).
- Robitaille, C. D. & Fauteux, D. Phase-diagrams and conductivity characterization of some PEO-Li_x electrolytes. *J. Electrochem. Soc.* **133**, 315–325 (1986).
- Grimaldi, J. J. & Lehn, J. M. Multicarrier transport: coupled transport of electrons and metal cations

mediated by an electron carrier and a selective cation carrier. *J. Am. Chem. Soc.* **101**, 1333–1334 (1979).

- Burla, M. C. *et al.* SIR88—a direct-methods program for the automatic solution of crystal-structures. *J. Appl. Crystallogr.* **22**, 389–393 (1989).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank N. Robertson for synthesis of some of the crystals used in this work, N. Nonose for measurements of the ICP mass spectra, and A. Yap and S. R. Elliott for assistance with the NMR measurements. This work was supported by Grant-in-aid for Science Research from the Ministry of Education, Science and Culture, Japan, and the UK Engineering and Physical Sciences Research Council. One of the authors (A.E.U.) acknowledges support from the British Council.

Correspondence and requests for materials should be addressed to R.H.F. (e-mail: rhf10@cam.ac.uk).

Magnitudes of sea-level lowstands of the past 500,000 years

E. J. Rohling*, M. Fenton*, F. J. Jorissen†, P. Bertrand‡, G. Ganssen‡ & J. P. Caulet§

* Department of Oceanography, Southampton University, Southampton Oceanography Centre, Southampton SO14 3ZH, UK

† Département de Géologie et Océanographie, Université de Bordeaux I, CNRS URA 197, Avenue des Facultés, 33405 Talence Cedex, France

‡ Department of Earth Sciences, Free University Amsterdam, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands

§ Laboratoire de Géologie, National Museum for Natural History, CNRS URA 723, 43 Rue Buffon, 75005 Paris, France

Existing techniques for estimating natural fluctuations of sea level and global ice-volume from the recent geological past exploit fossil coral-reef terraces or oxygen-isotope records from benthic foraminifera. Fossil reefs reveal the magnitude of sea-level peaks (highstands) of the past million years, but fail to produce significant values for minima (lowstands) before the Last Glacial Maximum (LGM) about 20,000 years ago, a time at which sea level was about 120 m lower than it is today^{1–4}. The isotope method provides a continuous sea-level record for the past 140,000 years (ref. 5) (calibrated with fossil-reef data⁶), but the realistic uncertainty in the sea-level estimates is around $\pm 20 \text{ m}$. Here we present improved lowstand estimates—extending the record back to 500,000 years before present—using an independent method based on combining evidence of extreme high-salinity conditions in the glacial Red Sea with a simple hydraulic control model of water flow through the Strait of Bab-el-Mandab, which links the Red Sea to the open ocean. We find that the world can glaciate more intensely than during the LGM by up to an additional 20-m lowering of global sea-level. Such a 20-m difference is equivalent to a change in global ice-volume of the order of today's Greenland and West Antarctic ice-sheets.

Our technique relies on evidence of adverse living conditions for planktonic foraminifera in the glacial Red Sea due to extremely high salinities, giving rise to so-called 'aplanctonic' zones. These contain only few specimens (mostly *Globigerinoides ruber*) and are found not only in our core MD921017 (Fig. 1), but throughout the Red Sea^{7–11}. Planktonic pteropods¹⁰ and benthic foraminiferal faunas were less affected, although the latter show increased abundances of high-salinity-resistant miliolid taxa (Fig. 1e; also refs 8, 9). Inorganic aragonite coatings, cements and concretions indicative of high salinities are found within the 'aplanctonic' zones of stages 2, 6 and 12 (Fig. 1b), in agreement with previous reports^{12,13}. Multi-disciplinary studies of the youngest aplanctonic zone (LGM) indicate that Red Sea salinities rose to 50 ± 2 practical salinity units (p.s.u.), that is, 10 ± 2 p.s.u. higher than in the adjacent ocean^{9,10,14}. The highest estimate¹¹ is 55 p.s.u. Similar faunal reductions and compositional changes, as well as distinct aragonite

occurrences, characterize the stage 6 and 12 glacial maxima, reflecting comparable high-salinity conditions (Fig. 1). Continuation of a more diverse, although severely reduced, fauna through stages 10 and 8 suggests less harsh conditions, with those during stage 8 being least restricted.

Arguments for hydraulic control of flow through the Strait of Bab-el-Mandab (SBM) and conservation of mass and salt^{14,15} demonstrate that the glacial freshwater deficit was similar to the present (2 myr^{-1}), and that the high salinities resulted from restricted marine exchange through the SBM, which today is only 137 m deep¹⁶. Continuation of benthic faunas, albeit in reduced numbers and different compositions, indicates that all glacial sea-level drops of the past 500 kyr left sufficient communication between the Red Sea and the open ocean to prevent worse salinization and consequent sterilization (Fig. 1c, e). The calculations^{14,15} may be rearranged to find the 'critical' glacial sill depth (H_{crit}) at which salinities would rise to reconstructed LGM values, and so trigger an aplanktonic interval. We use the more common range of estimates for LGM salinities^{9,10,14}, with a salinity difference (ΔS) across the SBM of $10 \pm 2 \text{ p.s.u.}$, and in sensitivity tests with $\Delta S = 15 \text{ p.s.u.}$ and $\Delta S = 10 \pm 5 \text{ p.s.u.}$ we evaluate the potential effects of the highest-salinity extreme¹¹.

A glacial water deficit of $2.0 \pm 0.5 \text{ myr}^{-1}$ is used, allowing for uncertainties in sea-air temperature contrasts and freshwater input of a factor of two, and $\pm 2 \text{ ms}^{-1}$ uncertainty in mean

wind speed variations¹⁴. As glacial exposure of shelves would reduce the Red Sea surface area by 50%, volumes of net glacial evaporation ($= \text{area} \times \text{deficit}$) amounted to 0.50 ± 0.13 times present-day values. Hydraulic control defines $Q = W0.375H_{\text{crit}}(0.375H_{\text{crit}}7.4 \times 10^{-3}\Delta S)^{0.5}$ for the maximum exchange solution, while $Q = (3.29\gamma Q_p)/\Delta S$ follows from conservation of mass and salt^{14,15}. Here Q is volume of Red Sea outflow, W is the glacial width of the shallow passage of the SBM, and γ is the ratio of glacial to present-day net evaporation; subscript p indicates present-day value.

Consequently, $H_{\text{crit}} = 18 \pm 5 \text{ m}$, for $\Delta S = 10 \pm 2 \text{ p.s.u.}$, $W = 11 \pm 1 \text{ km}$, $\gamma = 0.50 \pm 0.13$ and $Q_p = 0.32 \pm 0.03 \text{ Sv}$ ($1 \text{ Sv} = 10^6 \text{ m}^3 \text{ s}^{-1}$)¹⁴⁻¹⁷. Potentially increased mixing of inflow and outflow for reduced sill depths favours values of H_{crit} towards the higher end of its confidence interval over those towards the lower end. We combine the present-day sill depth of 137 m with H_{crit} to estimate past sea-level lowstands, but we first evaluate whether—and to what extent—correction is needed for uplift in the SBM region.

Total planktonic foraminiferal abundances show a distinct decrease over the interglacial periods of the past 500 kyr in core MD921017 (Figs 1b, 2a). A similar trend is seen for the past 380 kyr contained in nearby core KL11¹¹. Sedimentation rates throughout MD921017 are very uniform (see Supplementary Information), so that the foraminiferal numbers per gram—based on constant sample thickness—closely approximate true fluxes per unit time.

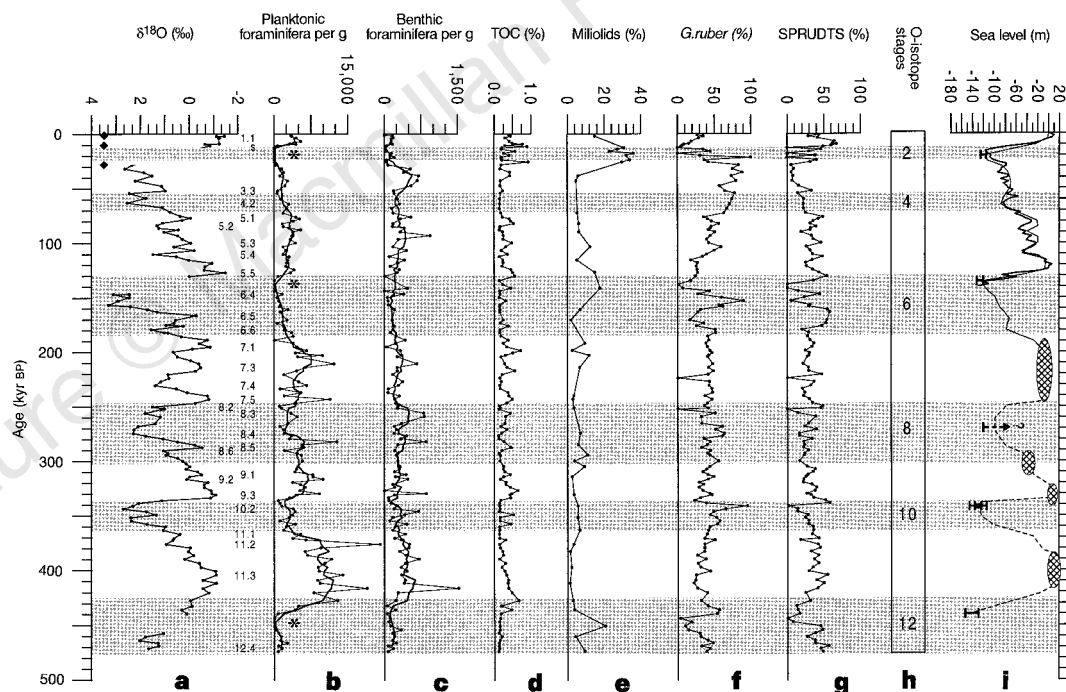


Figure 1 Results for core MD921017, Red Sea ($19^{\circ}23'24''\text{N}$, $38^{\circ}40'84''\text{E}$, water depth 570 m). All variables plotted against time in calendar years (kyr BP) according to correlation with SPECMAP²². **a**, Oxygen-isotope record (versus PDB standard) for *Globigerinoides ruber* in the 250–350 μm size fraction (filled circles). Numbers refer to SPECMAP events²²; filled diamonds indicate samples used for AMS radiocarbon dating ($1,666 \pm 26$, $9,681 \pm 41$ and $25,300 \pm 200$ radiocarbon years BP, respectively). **b**, Planktonic foraminifera per gram dry weight (thin line). Asterisks highlight true aplanktonic zones of oxygen isotope stages 2, 6 and 12 which contain aragonitic coatings, cement and nodules. Heavy line is five-point moving-average highlighting the general trends. **c**, Benthic foraminifera per gram dry weight (thin line), with five-point moving average (heavy line). We note factor 10 difference between scales in **b** and **c**. **d**, Down-core variation in TOC (%) (LECO 125-CS elemental analyser). Preparation involved removal of carbonate by acidification using dilute hydrochloric acid. **e**, Abundance of miliolids relative to total benthic foraminiferal fauna. Although of lower resolution than the other records, the miliolid percentages do show a general increase in abundance

through the record, as well as peaks associated with the aplanktonic zones. **f**, Abundance of *Globigerinoides ruber* relative to total planktonic foraminiferal fauna. **g**, Abundance of SPRUDTS group relative to total planktonic foraminifera. This group represents a cluster of the individually infrequent subtropical species *Globigerinoides sacculifer*, *Hastigerina pelagica*, *Globoturborotalia rubescens*, *Orbulina universa*, *Globigerina digitata*, *Globoturborotalia tenella* and *Globigerinella siphonifera*¹⁸. **h**, Oxygen isotope stages²². **i**, Global sea level record. The heavy solid line through the past 140 kyr is based on benthic isotopes⁵ and the thin solid line through the past 200 kyr is based on coral reef terraces and benthic isotopes⁴. Cross-hatched ovals show reported ranges of interglacial sea-level highstands^{2-4,21}. Thick error bars in stages 2, 6, 8, 10 and 12 represent ranges of glacial sea-level lowstands according to the model presented here. The dashed line through stages 8–12 shows schematic sea-level fluctuations sketched through the control points following the main trends in the oxygen-isotope record.

Consequently, the decreasing trend in planktonic foraminiferal numbers reflects: (1) a steady decrease in interglacial productivity over the past 500 kyr; or (2) a progressive deterioration of living conditions for foraminifera through increasing isolation of the Red Sea from the open ocean.

Investigation of other parameters for interglacial periods, to detect long-term changes unrelated to glacial cycles, provides no support for the first option. First, the total organic carbon (TOC) record shows only minor, random fluctuations (Fig. 1d). Second, there are no long-term trends in the carbon isotope record. Third,

the continuously low-diversity planktonic foraminiferal faunas are totally dominated by relative fluctuations (within a 100% sum) of only two taxonomic groups, *G. ruber* and the SPRUDTS group (see Fig. 1 legend for details), both of which are typical of warm and oligotrophic conditions¹⁸ (Figs 1f, g, 2b). Last, there is only a weak, statistically insignificant, trend in total benthic foraminiferal abundances, whereas a significant decrease would be expected under conditions of decreasing primary/export productivity (= food supply) (Figs 1c, 2a).

The second option—increasing Red Sea ‘restriction’—is supported by a steady increase in the abundances of high-salinity resistant^{8,9} miliolid species in the benthic foraminiferal fauna over the interglacial periods of the past 500 kyr (Figs 1e, 2c). The increasing restriction would have occurred within a context of already limited exchange with the open ocean, as witnessed by: the lack of successful¹⁹ ‘invasions’ into the Red Sea by oceanic species like *Globigerina bulloides*, *Globorotalia menardii* and *Neogloboquadrina dutertrei*; the limited and diminishing presence of *Turborotalita quinqueloba* (Fig. 2b) (these four species abound just outside the SBM in core MD921005; 11° 35′ 13″ N, 43° 31′ 84″ E); and the amplified glacial–interglacial contrasts throughout the Red Sea oxygen-isotope record relative to those of the open ocean (Fig. 1a), a typical characteristic of evaporative marginal basins related to their restriction from the world ocean. The inferred slow but progressive isolation of the Red Sea, superimposed on already restricted communication with the open ocean, is best explained by slow and fairly continuous uplift of a shallow sill.

The rate of uplift may be assessed using the present-day sill depth of 137 m (ref. 16) and previous estimates that sea level stood around 125 m below present sea level (b.p.s.l.) during stage 6 (135 kyr BP) (Fig. 1i)^{4,5}. Addition of H_{crit} gives an estimated stage 6 sill depth of 143 m b.p.s.l., which implies 6 m uplift in 135 kyr (4.4 cm kyr⁻¹). This rate is a factor of 5–10 smaller than the assumed constant rates of classical reef-terrace-based sea-level studies^{1–4}. To account for uncertainties in the derivation of the rate and/or its possible temporal nonlinearities, a 50% confidence interval is added, giving 0.044 ± 0.022 m kyr⁻¹. Our model then back-calculates the stage 6 sea level drop at $137 + (135 \times 0.044) - H_{crit} = 125$ m, with a confidence interval of ± 6 m based on propagation²⁰ of uncertainties in H_{crit} (± 5 m) and uplift rate (± 0.022 m kyr⁻¹). Validation of the model through determination of the LGM

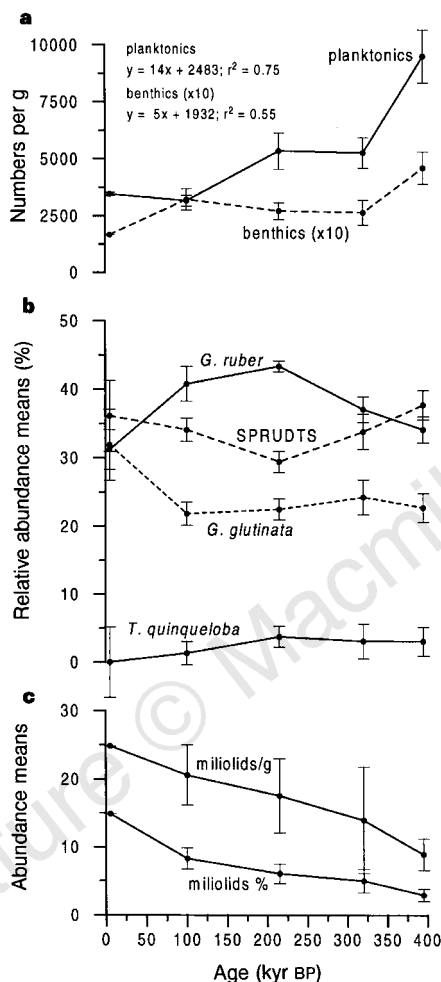


Figure 2 Trends through interglacial stages. **a**, Trends in the numbers of planktonic and benthic foraminifera per gram sediment dry weight, through interglacial stages 11, 9, 7, 5 and 1. Positions on age axis are simple stage mid-points. Stage 3 is excluded as this in fact is a warm interstadial within the last glacial period. Symbols indicate mean values, and error bars an interval of ± 1 standard error of the mean. Benthic numbers have been multiplied times 10, to allow plotting on the same scale as planktonic numbers. Equations and r^2 coefficients concern linear fits through the two records. Linear fit for planktonics shows a significant trend ($\alpha = 0.05$), for benthics the trend is weak and statistically insignificant. We emphasise that, as stage 1 is continuing, its values may be less representative than those of the other (completed) interglacial periods. **b**, Mean relative abundances of *Globigerinoides ruber* and SPRUDTS-group (see also Fig. 1), and of *Globigerinita glutinata* and *Turborotalita quinqueloba*, in percentages relative to total planktonic foraminiferal fauna, through interglacial stages 11, 9, 7, 5 and 1. Symbols, error bars and note on stage 3 and 1 values as in **a**. **c**, Trends in high-salinity indicative miliolid benthic foraminifera—both relative to total benthic foraminiferal fauna (%), and as absolute abundances (numbers per gram dry weight)—through interglacial stages 11, 9, 7, 5 and 1. Symbols, error bars and note on stage 3 and 1 values as in **a**.

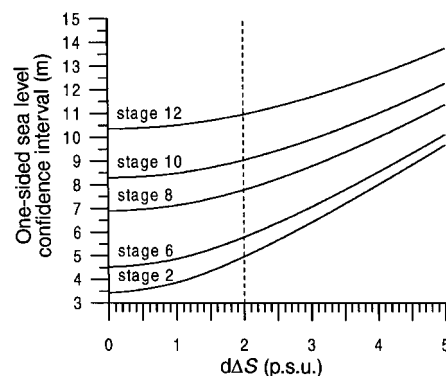


Figure 3 Sensitivity of confidence interval. Changes in the confidence intervals around the reconstructed mean sea level lowstands for glacial stages, 2, 6, 8, 10 and 12, for glacial Red Sea salinity contrast $\Delta S = 10$ p.s.u. $\pm$ a variable confidence interval $d\Delta S$. Effects are investigated per glacial stage for a range of $d\Delta S$ values from 0 to 5 p.s.u. Dashed line indicates $d\Delta S = 2$ as used in this Letter. The sea level confidence intervals (y axis) essentially change due to modification of the confidence interval around H_{crit} , which by close approximation corresponds to the line marked “stage 2”.

(20 kyr BP) sea-level lowstand gives 120 ± 5 m, in agreement with fossil reef results¹ and supporting our mean value for H_{crit} .

To evaluate the dependence of our lowstand reconstructions on the main assumption that previous glacial ΔS values were similar to LGM values of 10 ± 2 p.s.u., we perform two tests based on the highest extreme glacial salinity estimate (55 p.s.u.; ref. 11). One increases ΔS to 15 p.s.u.; the other keeps ΔS at 10 p.s.u. but increases its confidence interval ($d\Delta S$) from ± 2 to ± 5 p.s.u. In the first test, H_{crit} is smaller, and reconstructed sea level drops are greater, by a maximum of 6 m. In the second test, with $d\Delta S$ increased by a factor of 2.5, the lowstand confidence intervals remain accurate within a factor of 2 (Fig. 3). These results justify the use of the SBM uplift rate with H_{crit} to study pre-stage-6 lowstands, noting that our method is more likely to underestimate rather than overestimate past sea-level drops, by a few metres.

Accounting for uplift since stage 8 (270 kyr BP), sill depth was around 150 m b.p.s.l. Stage 8 does not contain a completely 'aplanktonic' interval. Although the total planktonic foraminiferal numbers are strongly reduced, the main species composition shows little change that might reflect high-salinity stress (Fig. 1). We infer that sill depth remained considerably greater than H_{crit} . To allow continuation of all observed planktonic species, Red Sea salinity should have remained below a maximum of ~ 45 p.s.u., requiring a minimum sill depth of ~ 30 m (compare ref. 14). Hence, the maximum conceived stage 8 sea level drop is $150 - 30 = 120$ m b.p.s.l. (± 8 m).

A similar argument to that for stage 8 may be made for stage 10 (340 kyr BP). However, stage 10 shows a much closer approximation of a complete 'aplanktonic' zone, with disruption of the main species composition. We infer that sill depth was maintained between H_{crit} (18 m) and 30 m, defining a stage 10 lowstand between 134 and 122 m b.p.s.l. (± 9 m).

Stage 12 (440 kyr BP) contains a true 'aplanktonic' zone (Fig. 1), suggesting a sill depth around H_{crit} and, consequently, a sea-level lowstand of 139 m b.p.s.l. (± 11 m). This mean value implies that global ice-volume during stage 12 exceeded LGM values by some 15%. This independently derived result validates the only previous estimate of stage 12 ice-volume, based on benthic oxygen-isotope records⁵.

Our lowstand values allow assessment of sea-level rises during the main deglaciations of the past 500 kyr (Fig. 1i), for comparison with that of 120 m following the LGM¹. With the maximum stage 5 sea level ~ 6 m above the present^{2,6}, the stage 6–5 sea-level rise was around 131 ± 6 m. During interglacial stage 7, sea level remained below the present-day level^{4,5}, giving a maximum amplitude for the stage 8–7 sea-level rise of 120 m, although the actual rise was probably considerably smaller. The stage 9 highstand reached 0–15 m above the present-day level^{4,5}, giving a stage 10–9 sea-level rise between 122 and 149 m. The largest sea-level rise of the past 500 kyr followed the stage 12 lowstand of 139 ± 11 m b.p.s.l. and culminated in a maximum stage 11 highstand up to 20 m above present-day sea level²¹.

We conclude that the last glacial–interglacial cycle showed ice-volume fluctuations that were more than 10% smaller than those that occurred in three out of four of the immediately preceding main cycles. The stage 12–11 sea level rise implies that over 30% greater ice-volume changes were involved in Quaternary glacial–interglacial cycles than would be expected on the basis of the last cycle alone. □

Received 27 October 1997; accepted 22 April 1998.

1. Fairbanks, R. G. A 17,000 year glacio-eustatic sea level record: Influence of glacial melting rates on the Younger-Dryas event and deep ocean circulation. *Nature* **342**, 637–642 (1989).
2. Radtke, U. & Grün, R. Revised reconstruction of middle and late Pleistocene sea-level changes based on new chronologic and morphologic investigations in Barbados, West Indies. *J. Coastal Res.* **6**, 699–708 (1990).
3. Pirazzoli, P. A. et al. A one million-year-long sequence of marine terraces on Sumba Island, Indonesia. *Mar. Geol.* **109**, 221–236 (1993).
4. Bard, E. et al. Pleistocene sea levels and tectonic uplift based on dating of corals from Sumba Island, Indonesia. *Geophys. Res. Lett.* **23**, 1473–1476 (1996).

5. Shackleton, N. J. Oxygen isotopes, ice volume and sea level. *Quat. Sci. Rev.* **6**, 183–190 (1987).
6. Chappell, J. & Shackleton, N. J. Oxygen isotopes and sea level. *Nature* **324**, 137–140 (1986).
7. Berggren, W. A. & Boersma, A. in *Hot Brines and Heavy Metal Deposits* (eds Degens, E. T. & Ross, D. A.) 282–298 (Springer, New York, 1969).
8. Halicz, E. & Reiss, Z. Palaeoecological relations of foraminifera in a desert enclosed sea—The Gulf of Aqaba. *Mar. Ecol.* **2**, 15–34 (1981).
9. Locke, S. & Thunell, R. C. Palaeoceanographic record of the last glacial–interglacial cycle in the Red Sea and Gulf of Aden. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **64**, 163–187 (1987).
10. Almogi-Labin, A., Hemleben, C., Meischner, D. & Erlenkeuser, H. Palaeoenvironmental events during the last 13,000 years in the central Red Sea as recorded by pteropoda. *Paleoceanography* **6**, 83–98 (1991).
11. Hemleben, C. et al. Three hundred and eighty thousand year long stable isotope and faunal records from the Red Sea: Influence of global sea level change on hydrography. *Paleoceanography* **11**, 147–156 (1996).
12. Milliman, J. D., Ross, D. A. & Ku, T. L. in *Hot Brines and Heavy Metal Deposits* (eds Degens, E. T. & Ross, D. A.) 724–736 (Springer, New York, 1969).
13. Ku, T. L., Thurber, D. L. & Mathieu, G. G. in *Hot Brines and Heavy Metal Deposits* (eds Degens, E. T. & Ross, D. A.) 348–359 (Springer, New York, 1969).
14. Rohling, E. J. Glacial conditions in the Red Sea. *Paleoceanography* **9**, 653–660 (1994).
15. Rohling, E. J. & Zachariasse, W. J. Red Sea outflow during the last glacial maximum. *Quat. Int.* **31**, 77–83 (1996).
16. Werner, F. & Lange, K. A bathymetric survey of the sill area between the Red Sea and Gulf of Aden. *Geol. Jahrb. D* **13**, 125–130 (1975).
17. Siedler, G. in *Hot Brines and Heavy Metal Deposits* (eds Degens, E. T. & Ross, D. A.) 131–137 (Springer, New York, 1969).
18. Rohling, E. J., Jorissen, F. J., Vergnaud-Grazzini, C. & Zachariasse, W. J. Northern Levantine and Adriatic Quaternary planktic foraminifera; Reconstruction of paleoenvironmental gradients. *Mar. Micropaleontol.* **21**, 191–218 (1993).
19. Ganssen, G. & Kroon, D. Evidence for Red Sea surface circulation from oxygen isotopes of modern surface waters and planktonic foraminiferal tests. *Paleoceanography* **6**, 73–82 (1991).
20. Squires, G. L. *Practical Physics* 3rd edn (Cambridge Univ. Press, 1988).
21. Howard, W. R. A warm future in the past. *Nature* **388**, 418–419 (1997).
22. Imbrie, J. et al. in *Milankovitch and Climate* (eds Berger, A. et al.) 269–305 (Reidel, Hingham, MA, 1984).

Supplementary Information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank H. Vonhof, M. Dignan and P. Martinez for assistance with stable-isotope and TOC analyses; J. W. Zachariasse for cooperation within the context of our joint studies of the NW Indian Ocean; NERC for support to M.F., and the National Museum of Natural History in Paris for support to E.J.R. during the planning and sampling phase of this work.

Correspondence and requests for materials should be addressed to E.J.R. (e-mail: E.Rohling@soc.soton.ac.uk).

Megaripple migration in a natural surf zone

Edith L. Gallagher*, Steve Elgar† & Edward B. Thornton*

* Oceanography Department, Naval Postgraduate School, Monterey, California 93943, USA

† School of Electrical Engineering and Computer Science, Washington State University, Pullman, Washington 99164, USA

Migrating megaripples are bedforms that appear in the surf zone of sandy coasts¹. With heights of 0.1–0.5 m and wavelengths of 1–5 m, they are similar in size and shape to small dunes, large ripples, or sand waves. Such sedimentary bedforms have been studied in subaerial², steady-flow³ and intertidal⁴ environments, as well as in laboratory flume experiments⁵. They affect overlying currents by introducing hydraulic roughness^{4,6}, and may provide a mechanism for sediment transport^{7,8} as well as forming sedimentary structures in preserved facies^{9,10}. The formation, orientation and migration of such bedforms is not understood well^{11,12}. Dunes, for example, can be aligned with their crests perpendicular to steady unidirectional winds¹³, but in more complex wind fields their orientation becomes difficult to predict^{14–17}. Similarly, it is not known how sea-floor megaripples become aligned and migrate in the complex flows of the surf zone. Here we present observations in the surf zone of a natural beach which indicate that megaripples do not migrate in the direction of the vector sum of the currents, but are aligned so that the sediment transport normal to the bedform crest is maximized¹⁷. This may need to be taken into account in modelling morphology change and interpreting existing and fossil morphologic patterns.

Lunate-shaped megaripples migrating shoreward have been observed by scuba divers in a high-energy surf zone on the Oregon coast¹ and with sonar images of the sand bed in the surf zone of a large lake during the waning stages of a storm¹⁸. Onshore migration has been proposed to result from sediment transport associated with asymmetries in wave velocities¹⁹, but migration with steady unidirectional flows (rip, longshore and tidal currents) has also been observed^{20–23}. We have made observations in the surf zone of a natural beach (Fig. 1a) with an array of seven downward-looking sonar altimeters²⁴ that measure the distance to the sea floor from a fixed frame (Fig. 1b). Megaripples (often visually determined to be slightly lunate shaped) were present about 60% of the time during the six-week experiment, which included a wide range of wave heights (from 0.1 to 4.0 m) and mean currents (from 0 to 2 m s⁻¹). An example of migrating megaripples is shown in Fig. 2.

To determine migration rate and direction quantitatively, we used time domain cross correlation to calculate the time lag (Δt) of maximum correlation between sensor pairs, separated by distances Δx and Δy in the cross- and alongshore directions, respectively. The

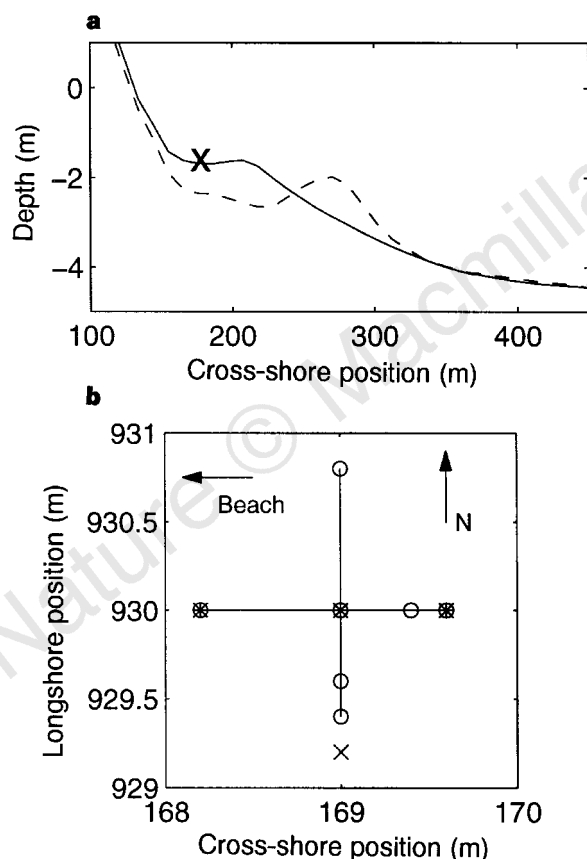


Figure 1 Location of the altimeter array in the surf zone. **a**, Depth of the sea floor (relative to mean sea level) versus cross-shore position. Profiles from the start (1 Sep, solid curve) and end (14 Oct, dashed curve) of the observation period are shown. The array was located ~50 m from the shoreline in the trough (X) between the sandbar and the beach on a barrier island exposed to the Atlantic Ocean near the town of Duck, North Carolina, USA. The mean water depth at the array ranged from 1.5 to 2 m between 1 September and 14 October 1994 as the large-scale beach morphology evolved. The amplitude of the semidiurnal tide near Duck is about 0.5 m and the mean sand-grain diameter near the array was ~0.2 mm. **b**, Plan view of the altimeter array. Circles represent locations of individual altimeters (time series from those with asterisks are shown in Fig. 2) and the cross shows the location of an electromagnetic current meter that measured cross- and alongshore components of the fluid velocity field every 0.5 s about 0.5 m above the sea floor.

components of megaripple migration velocity ($\Delta x/\Delta t$, $\Delta y/\Delta t$) calculated from each sensor pair (when the maximum correlation was greater than 0.4) were averaged, giving an estimate of cross- (U_{Mrip}) and along shore (V_{Mrip}) migration speed and direction $\theta_{\text{Mrip}} = \text{atan}(V_{\text{Mrip}}/U_{\text{Mrip}})$. Average (over 48 h) migration speeds range from 0.1 to 1.7 m h⁻¹.

The measured cross- and alongshore fluid velocities (Fig. 1) were separated into mean and fluctuating (that is, wave) components and 48-h mean cross- ($\bar{U}_{48}$) and alongshore ($\bar{V}_{48}$) velocities were calculated. The direction of the 48-h mean current $\theta_{\text{mean}} = \text{atan}(\bar{V}_{48}/\bar{U}_{48})$ was rarely aligned with the direction of megaripple migration, θ_{Mrip} (Fig. 3a). Megaripples were aligned with the alongshore ($\sim 90^\circ$ in Fig. 3a) steady flow ($\bar{V}_{48}$) only during periods when $\bar{V}_{48}$ was relatively large (more than about 1.5 times the root-mean-square (r.m.s.) wave velocity and $\bar{V}_{48} \gg \bar{U}_{48}$).

Hourly skewness-weighted r.m.s. cross- ($\langle \bar{u}^3 \rangle / \langle \bar{u}^2 \rangle$) and alongshore ($\langle \bar{v}^3 \rangle / \langle \bar{v}^2 \rangle$) wave velocities, where $\bar{u}$, $\bar{v}$ are the cross- and alongshore components of the fluctuating velocities, respectively and $\langle \rangle$ denote time average, were averaged to give 48-h cross- ($\bar{U}_{48}$) and alongshore ($\bar{V}_{48}$) wave velocities with magnitude $\sqrt{\bar{U}_{48}^2 + \bar{V}_{48}^2}$ and direction $\theta_{\text{wave}} = \text{atan}(\bar{V}_{48}/\bar{U}_{48})$. The skewness-weighted r.m.s. is used because it represents the magnitude of the wave velocities, similar to the r.m.s. velocity, as well as the direction of the onshore skewed waves in the surf zone, believed to be important to net sediment transport²⁵. Megaripples were most closely aligned with the onshore ($\sim 0^\circ$ in Fig. 3b) directed wave velocities (consistent with previous visual observations^{1,18,19}) when the mean currents were relatively small. However, in general the r.m.s. difference between θ_{Mrip} and θ_{wave} is large and θ_{wave} cannot be used to predict θ_{Mrip} (Fig. 3b).

Megaripple migration direction also is not correlated with the direction of the vector sum of the mean and wave velocities $\theta_{\text{total}} = \text{atan}((\bar{U}_{48} + \bar{U}_{48})/(\bar{V}_{48} + \bar{V}_{48}))$ (Fig. 3c). Comparison of the migration directions with θ_{total} (Fig. 3c) indicates that these observations include transverse, longitudinal and oblique bedforms²⁶.

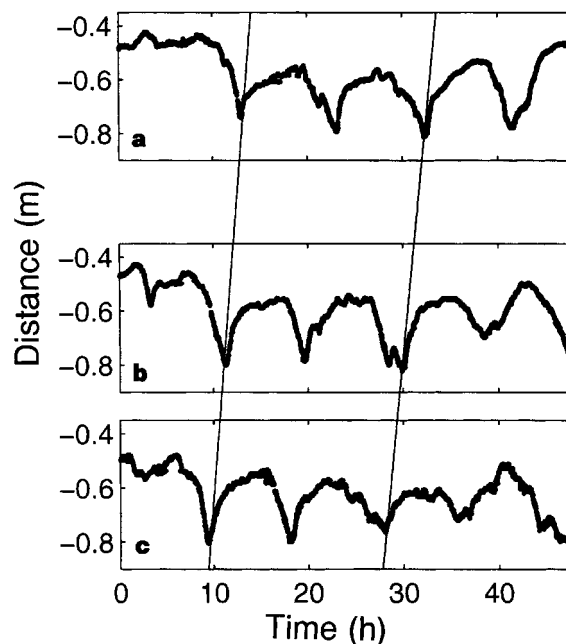


Figure 2 Distance to the sea floor (every 32 s (ref. 24) for 48 h) from three altimeters (circled asterisks in Fig. 1b) versus time. The altimeters were separated in the cross-shore direction by 80 (upper (a) to middle (b) panel) and 60 cm (middle (b) to lower (c) panel) and the panels are separated accordingly. Time series in the top panel are from the most onshore altimeter. The slopes of the lines connecting the 'troughs' of the bedforms illustrate onshore migration of the features (~ 0.3 m h⁻¹).

A similar lack of alignment of subaerial bedforms with the resultant transport vector in quasi-steady but directionally changing flows has been observed previously^{17,27}. Experiments in which a large sand-covered table was rotated periodically in steady wind led to the hypothesis that bedforms are not aligned with the resultant transport vector $\mathbf{R}$, but so that the gross sediment transport $T = |T_s| + |T_D|$ (Fig. 4a) normal to their crests is maximized¹⁷. Using gross transport allows back-and-forth motion of the sand to build a feature even if the net transport across the bedform is zero. For bidirectional flows, the orientation α of the bedform, relative to

the dominant transport vector $\mathbf{D}$, which maximizes the sum of the absolute value of bedform-normal components of the transport vectors $T = |\mathbf{D}|\sin\alpha + |\mathbf{S}|\sin(\gamma - \alpha)$, where $\mathbf{S}$ is the subordinate transport vector at angle γ relative to $\mathbf{D}$ (Fig. 4a), is given by¹⁷

$$\tan\alpha = \pm \frac{|\mathbf{D}|/|\mathbf{S}| + |\cos\gamma|}{|\sin\gamma|} \quad (1)$$

Equation (1) is applicable when the durations of the two temporally distinct transport vectors ($\mathbf{D}$ and $\mathbf{S}$) are each short enough that the bedforms do not come into equilibrium with either prevailing flow²⁸. In the mixed flows on a natural beach, mean and wave velocities act simultaneously and thus cannot be used to represent $\mathbf{D}$ and $\mathbf{S}$. However, the opposing oscillatory flows, modified in amplitude and direction by mean currents, are temporally distinct. The periods (order 10 s) of the wave velocities are far shorter than the time required for megaripples to equilibrate, so equation (1) is valid.

In a natural surf zone, there is large range in the magnitude and direction of individual waves, and thus the bidirectional solution

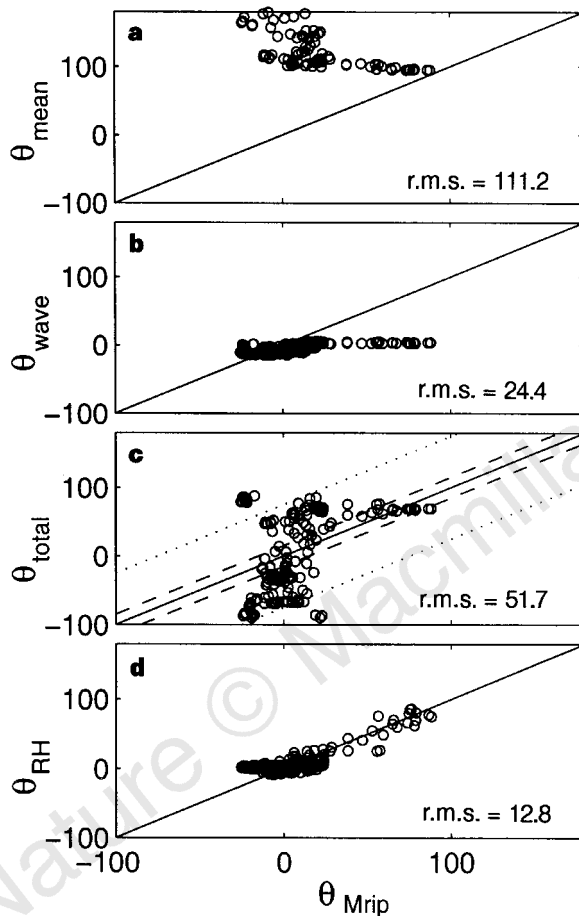


Figure 3 Direction of fluid flow versus megaripple migration. Direction of **a**, mean current; **b**, wave velocity; **c**, vector sum of mean and wave velocities; and **d**, bedform orientation (θ_{RH}) predicted by maximizing the gross, bedform-normal transport T (equation (2))¹⁷ versus observed megaripple migration direction (in degrees with 0° onshore and 90° to the south). All values are 48-h averages, which include enough migrating features for statistical stability, while retaining stationarity of the slowly evolving bedforms. The features change constantly, so overlapping 48-h estimates starting every 3 h were used. Changing the size of the averaging period or the overlap by a factor of 2 does not strongly affect the results presented here. The root-mean-square (r.m.s.) difference between the dependent variable and the observed megaripple migration direction (θ_{Mrip}) is listed in each panel. The r.m.s. differences between θ_{RH} and θ_{Mrip} (12.8) are smaller than those between θ_{Mrip} and the directions of the mean current (111.2), the wave-orbital velocities (24.4), or their resultant (51.7). In **c**, the broken lines delineate regions occupied by traditional bedform classes²⁶. Transverse bedforms ($0^\circ \leq |\theta_{total} - \theta_{Mrip}| < 15^\circ$) fall inside the dashed lines, longitudinal bedforms ($75^\circ < |\theta_{total} - \theta_{Mrip}| \leq 90^\circ$) fall outside the dotted lines and oblique bedforms ($75^\circ < |\theta_{total} - \theta_{Mrip}| \leq 75^\circ$) fall in between. The correlation between θ_{RH} and θ_{Mrip} (**d**) is $r = 0.86$, and changing the exponent of the velocity (that is, the model for sediment transport) from three to one or six results in correlations of $r = 0.85$ and $r = 0.87$, respectively. None is statistically different at the 99% level.

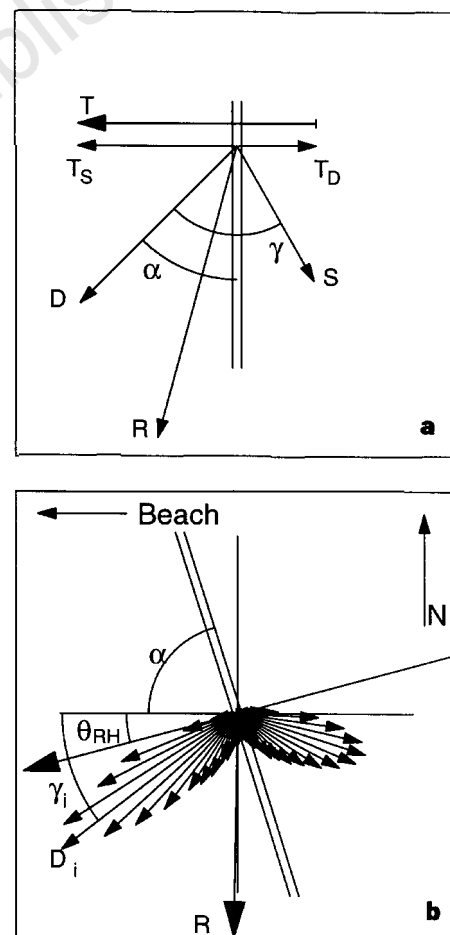


Figure 4 Transport vectors and bedform orientation¹⁷. **a**, Following ref. 17, the angle between the dominant ($\mathbf{D}$) and subordinate ($\mathbf{S}$) transport vectors is γ , the angle between $\mathbf{D}$ and the bedform crest (double line) is α , $\mathbf{R}$ is the vector sum of $\mathbf{D}$ and $\mathbf{S}$, and $T = |T_s| + |T_D|$ is the gross bedform-normal transport. **b**, Example of the directional distribution of cumulative transport (summed over a 3-h period) from instantaneous velocities in the surf zone (small arrows). The magnitude and direction of each cumulative transport vector are given by D_i and γ_i , respectively. The orientation of the bedform crest (double line) α is such that the gross bedform-normal transport T (equation (2)) is maximized. $\mathbf{R}$ shows the direction of the resultant vector for this example. Observed megaripple migration directions θ_{Mrip} are compared with $\theta_{RH} = 90^\circ - \alpha$ in Fig. 3d.

(equation (1)) is modified to a sum over a continuum of instantaneous transport vectors, with the component of gross transport normal to the bedforms given by

$$T = \sum_i D_i |\sin(\gamma_i - \alpha)| \quad (2)$$

where i ranges over all transport vectors. Although in this more general case of many transport vectors there is no analytical solution for α , the gross, bedform-normal transport T can be maximized numerically to give bedform orientation.

In the bidirectional, rotating table experiments¹⁷, transport vectors $\mathbf{D}$ and $\mathbf{S}$ were proportional to the duration of steady winds in each direction, and thus the bedform orientation α depended on the ratio $|\mathbf{D}|/|\mathbf{S}|$ (equation (1)) and not on a particular sediment transport model. However, the more general solution requires a sediment transport model to represent D_i . Similar to bedload models²³, D_i was calculated as proportional to the instantaneous velocity cubed. The results are not sensitive to the exponent of the velocity used to calculate D_i . Here, the individual instantaneous (2 Hz) transport vectors (calculated from velocity measurements) were summed over a 3-h period and sorted into 5°-wide directional bins, giving a directional distribution of cumulative transport (see, for example, Fig. 4b). The distribution was used to calculate T (equation (2)), with γ_i and D_i equal to the direction and corresponding magnitude of the cumulative transport in each bin. The value of α for which T is maximum was then found for each 3-h period. The model is sensitive to the estimation period and better results were obtained with shorter periods, implying that the bedforms respond quickly (order 3 h) to changes in the flow field.

For comparison with megaripple migration direction observations, 16 sequential 3-h estimates of α were averaged to give a 48-h estimate. Assuming that megaripples migrate $\sim 90^\circ$ to their crest orientation, a predicted migration direction is given by $\theta_{RH} = 90^\circ - \alpha$. As shown in Fig. 3d, θ_{RH} predicts accurately the observed megaripple migration direction θ_{Mrip} (the slope (1.04) of a best-fit line does not differ significantly from 1.00 at the 99.5% level).

These observations suggest that megaripple migration in the surf zone is caused by both mean and wave flows. However, the migration direction is not aligned with the vector sum of the currents, but so that gross sediment transport normal to the bedform is maximized, as suggested previously for subaerial features¹⁷. Megaripples in the surf zone of a natural barred beach occur frequently for a wide range of wave and current conditions, and include transverse, longitudinal and oblique bedforms. If migration of bedforms is an important mechanism for bedload sediment transport, parametrizations that depend only on waves, currents, or even their vector sum, may not predict the observed transport accurately. Conversely, flow conditions inferred from alignment of bed features preserved in ancient sedimentary deposits or observed in modern environments may not be unique because different flow fields can maximize gross bedform-normal transport. □

Received 29 December 1997; accepted 28 April 1998.

- Clifton, H. E., Hunter, R. E. & Phillips, R. L. Depositional structures and processes in the non-barred high-energy nearshore. *J. Sedim. Petrol.* **41**, 651–670 (1971).
- Bagnold, R. A. *The Physics of Wind Blown Sand and Desert Dunes* (Methuen, London, 1954).
- Simons, D. B. & Richardson, E. V. Forms of bed roughness in alluvial channels. *Am. Soc. Civ. Eng. Trans.* **128**, 284–302 (1963).
- Dalrymple, R. W., Knight, R. J. & Lambiase, J. J. Bedforms and their hydraulic stability relationships in a tidal environment, Bay of Fundy, Canada. *Nature* **275**, 100–104 (1978).
- Southard, J. B., Lambiase, J. M., Federico, D. C., Pile, H. T. & Weidman, C. R. Experiments on bed configurations in fine sands under bidirectional purely oscillatory flow, and the origin of hummocky cross-stratification. *J. Sedim. Petrol.* **60**, 1–17 (1990).
- Garcez Faria, A. F., Thornton, E. B., Stanton, T. P., Soares, C. V. & Lippmann, T. C. Vertical profiles of longshore currents and related bed shear stress and bottom roughness. *J. Geophys. Res.* **103**, 3217–3232 (1998).
- Allen, J. R. L. *Current Ripples: Their Relation to Patterns of Water and Sediment Motion* (North-Holland, Amsterdam, 1968).
- Middleton, G. V. & Southard, J. B. *Mechanics of Sediment Movement*. (Lecture for short course no. 3, Society of Economic Paleontologists and Mineralogists, 1982).
- Davidson-Arnott, R. G. D. & Greenwood, B. Facies relationships on a barred coast, Kouchibouguac

- Bay, New Brunswick, Canada. *Beach and Nearshore Sedimentation*, SEPM Special Pub. **24**, 149–169 (1976).
- Sherman, D. J. & Greenwood, B. Hummocky cross-stratification and post-vertex ripples: length scales and hydraulic analysis. *Sedimentology* **36**, 981–986 (1989).
- Werner, B. T. Eolian dunes: computer simulations and attractor interpretation. *Geology* **23**, 1107–1110 (1995).
- Stam, J. M. T. On the modelling of two-dimensional aeolian dunes. *Sedimentology* **44**, 127–141 (1997).
- Bagnold, R. A. Journeys in the Libyan Desert 1929 and 1939. *Geogr. J.* **78**, 13–39 (1933).
- Warren, A. Dune trend and the Ekman spiral. *Nature* **259**, 653–654 (1976).
- Tsoar, H. Dynamic processes operating on a longitudinal (seif) sand dune. *Sedimentology* **30**, 567–578 (1983).
- Hanna, S. R. The formation of longitudinal sand dunes by large helical eddies in the atmosphere. *J. Appl. Meteorol.* **8**, 874–883 (1969).
- Rubin, D. M. & Hunter, R. E. Bedform alignment in directionally varying flows. *Science* **237**, 276–278 (1987).
- Hay, A. E. & Wilson, D. J. Rotary sidescan images of nearshore bedform evolution during a storm. *Mar. Geol.* **119**, 57–65 (1994).
- Clifton, H. E. Wave-formed sedimentary structures—a conceptual model. *Beach and Nearshore Sedimentation*, SEPM Special Pub. **24**, 126–148 (1976).
- Moore, J. N., Fritz, W. J. & Futch, R. S. Occurrence of megaripples in a ridge and runnel system, Sapelo Island, Georgia: morphology and processes. *J. Sedim. Petrol.* **54**, 615–625 (1984).
- Sherman, D. J., Short, A. D. & Takeda, I. Sediment mixing depth and bedform migration in rip channels. *J. Coastal Res.* **15**, 39–48 (1993).
- Dalrymple, R. W. & Rhodes, R. N. Estuarine dunes and barforms. *Dev. Sedimentol.* **53**, 359–422 (1995).
- Hay, A. E. & Bowen, A. J. Alongshore migration of lunate megaripples during DUCK94. *J. Geophys. Res.* (in the press).
- Gallagher, E. L., Boyd, W., Elgar, S., Guza, R. T. & Woodward, B. Performance of a sonar altimeter in the nearshore. *Mar. Geol.* **133**, 241–248 (1996).
- Bagnold, R. A. An approach to the sediment transport problem from general physics. *US Geol. Surv. Prof. Paper* **422-I** (1966).
- Rubin, D. M. Cross-bedding, bedforms, and paleocurrents: Concepts in sedimentology and paleontology. *SEPM Special Pub.* **1** (1987).
- Fryberger, S. G. Dune forms and wind regime. *A Study of Global Sand Seas*, Pap. US Geol. Surv. **1052**, 137–170 (1979).
- Rubin, D. M. & Ikeda, H. Flume experiments on the alignment of transverse, oblique, and longitudinal dunes in directionally varying flows. *Sedimentology* **37**, 673–684 (1990).

Acknowledgements. We thank T. Drake for introducing us to ref. 17 and for sharing his insight into, and enthusiasm for, bedforms; R. T. Guza and B. T. Werner for suggestions and advice; D. Rubin for critically reading the manuscript; P. Fields for comments; and R. T. Guza, T. H. C. Herbers and B. Raubenheimer for help with acquiring field data. The instruments were fabricated, deployed and maintained by staff from the Center for Coastal Studies, Scripps Institution of Oceanography; the US Army Corps of Engineers Field Research Facility provided logistical support. Funding was from the Office of Naval Research, the NSF, and the National Research Council.

Correspondence should be addressed to either E.L.G. (email: egallagh@oc.nps.navy.mil) or S.E. (email: elgar@eecs.wsu.edu).

The unique anisotropy of the Pacific upper mantle

Göran Ekström & Adam M. Dziewonski

Department of Earth and Planetary Sciences, Harvard University,
20 Oxford Street, Cambridge, Massachusetts 02138, USA

The development and interpretation of tomographic models of the Earth's mantle have usually proceeded under the assumption that fast and slow seismic velocity anomalies represent a spatially heterogeneous temperature field associated with mantle convection. Implicit in this approach is an assumption that either the effect of anisotropy on seismic velocities is small in comparison with isotropic thermal or compositional effects, or that the tomographic results represent the average isotropic heterogeneity, even if individual seismic observations are affected by anisotropic structure. For example, velocity anomalies in the upper portions of the oceanic mantle are commonly interpreted in terms of the progressive cooling^{1,2} (and localized reheating³) of a mechanical and thermal boundary layer consisting of rigid oceanic lithosphere and an underlying, less viscous, asthenosphere. Here, however, we present results from a global three-dimensional tomographic model of shear-wave velocity which shows that the uppermost mantle beneath the central Pacific Ocean is considerably more complicated than this simple model. Over a broad area, with its centre near Hawaii, the seismic data reveal a regional anomaly in elastic anisotropy which produces variations of seismic velocities that are at least as large as

those due to thermal effects. Because seismic anisotropy is an indicator of strain in Earth materials, our tomographic results can be used to put constraints on both buoyancy forces (thermal effects) and flow patterns in the upper mantle.

Our finding comes in the context of the development of a high-resolution (up to spherical harmonic degree 20) global, three-dimensional (3D) tomographic model of shear-wave velocity in the Earth's upper mantle. Previously, using a large data set, we obtained an isotropic S-wave velocity model (S20-95; ref. 4) which showed some features different from those seen in the earlier tomographic mantle model S12 (ref. 5). The most striking disagreement between the models occurred at ~ 150 km depth under the Pacific plate. At this depth, the East Pacific Rise (EPR) is the most pronounced negative (slow) anomaly in model S12, as well as in other recent mantle models^{6,7}, whereas model S20-95 showed equally slow anomalies in the central Pacific. If model S20-95 were correct, this would have important implications for the controversy surrounding the depth extent of the velocity anomalies at mid-ocean ridges^{1,2}. At the same time, the variability of the anomalies shown by different tomographic models at shallow depths beneath the central Pacific would also cast doubt on the robustness of seismic tomographic results in general.

A number of experiments have led us to the conclusion that the difference seen in the Pacific between models S12 and S20-95, and possibly those seen between other tomographic models as well, is largely the result of incompatibility between two groups of data: one consisting of observations with primary sensitivity to V_{SH} (for example, Love waves), the shear-wave velocity of a transversely

polarized horizontally travelling S wave, and the other of observations sensitive to V_{SV} (for example, Rayleigh waves), the velocity of a vertically travelling S wave. In an isotropic material, V_{SH} and V_{SV} are equal, but in an anisotropic material they differ. The incorporation or exclusion, or relative weighting, of these different data in an inversion for isotropic S-wave velocity perturbations can lead to undesired compromises in the resulting models. For example, in the derivation of model S12, significantly higher weights were assigned to the V_{SH} -sensitive data, whereas S20-95 was dominated by V_{SV} -sensitive observations. The sensitivity of the results to such weighting schemes indicated to us a need for greater flexibility in the parametrization of the models, and motivated us to look more closely at the possibility of reconciling the different observations by consideration of regional variations in anisotropy.

Most published 3D global models of upper-mantle shear-wave velocity are isotropic, or consist of isotropic perturbations with respect to an anisotropic starting model. There are exceptions; Nataf *et al.*^{8,9}, for example, proposed global 3D models for V_{SV} and V_{SH} , thus implying spatially variable radial anisotropy (transverse isotropy with a radial symmetry axis). However, these models showed very large (up to 30%) differences in $|V_{SH} - V_{SV}|/V_{SH}$, a finding not supported by later studies. Montagner and Tanimoto^{10,11} investigated a more general form of mantle anisotropy, which included azimuthally varying terms, and found larger values for radial anisotropy in the oceanic mantle as compared with the continental mantle. Models derived for specific paths or regions have in several studies included radial anisotropy as an important component¹²⁻¹⁶, but often variations in anisotropy have been ignored, reflecting the

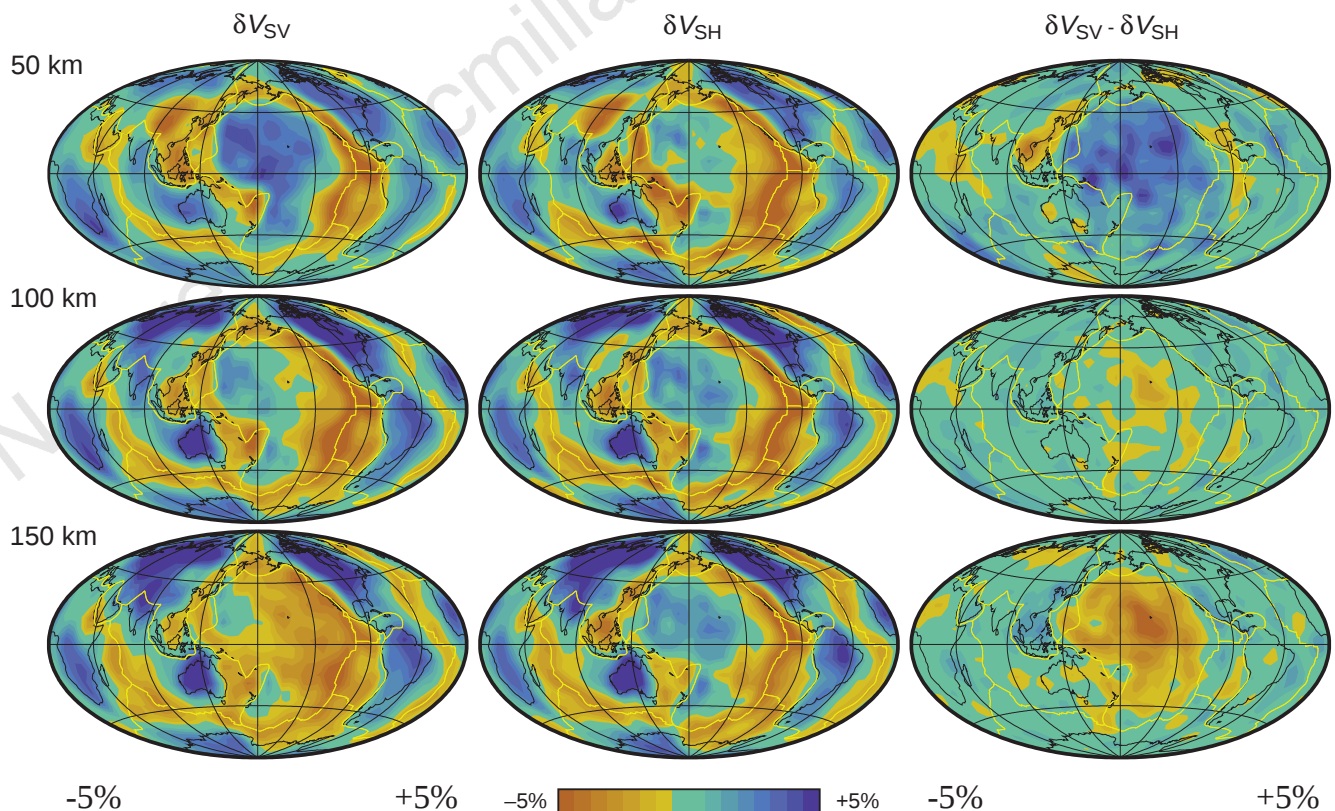


Figure 1 Maps showing the velocity variations in the model S20A at 50, 100, and 150 km depth. The left column shows the lateral variations in V_{SV} with respect to V_{SV}^{PREM} , $\delta V_{SV} = (V_{SV} - V_{SV}^{PREM})/V_{SV}^{PREM}$, the middle column the variations in V_{SH} with respect to V_{SH}^{PREM} , $\delta V_{SH} = (V_{SH} - V_{SH}^{PREM})/V_{SH}^{PREM}$ and the rightmost column the variations in their differences, $(\delta V_{SV} - \delta V_{SH})$. See Fig. 2 for examples of absolute variations in velocity. The top 200 km of the mantle is very well constrained by our data, primarily by the dispersion measurements of intermediate-period Love and Rayleigh waves. The most prominent velocity anomalies in this depth range are

associated with the fast mantle roots beneath continental interiors, the slow mid-ocean ridges and back-arc spreading centres, and the fast mantle beneath the oldest sea floor. The data sets used in the derivation of the V_{SH} and V_{SV} heterogeneities in the upper mantle are entirely independent. The similarity between the structures imaged by the two data sets beneath the continents is remarkable, and lends credence to the pronounced differences seen beneath the Pacific plate at 50 and 150 km depth.

assumption that anisotropic effects are small. Nishimura and Forsyth^{17,18} studied the variations in anisotropic properties of the Pacific upper mantle in detail—however, their analysis emphasized an age dependence for both isotropic and anisotropic properties. The results presented here suggest that this emphasis may have been too limiting.

Our new model S20A was derived using a wide variety of seismic observations and allowing for anisotropic heterogeneity in the upper mantle (see Methods for details). Figure 1 shows the perturbations in V_{SV} and V_{SH} with respect to their values in the reference model PREM¹⁹ at 50, 100 and 150 km depth in the mantle. At 50 km depth, the V_{SV} and V_{SH} models have very similar patterns of high velocities under the continents and low velocities under the mid-ocean ridges. A pronounced difference between the models is seen beneath the older portions of the Pacific plate. Here, the perturbation in V_{SV} is positive and large, while V_{SH} remains close to the PREM value. In comparison, the difference map for the rest of the world shows minor anomalies not associated with identifiable tectonic elements; such differences, smaller than about $\pm 1\%$, we consider to represent an unresolved background level.

At 100 km, the differences between the V_{SH} and V_{SV} models are small. The implication is that the radial anisotropy built into PREM, with $V_{SH} \sim 3\%$ faster than V_{SV} , is a good average for both continents and oceans. However, at 150 km depth, the pattern is dramatically different. Whereas V_{SH} shows the central Pacific as faster than the

global average, V_{SV} is significantly slower. The anomaly associated with the EPR is much better defined in V_{SH} , where it is clearly centred on the ridge axis; the negative V_{SV} anomalies are instead diffuse, extending to the centre of the Pacific. The maximum difference between the V_{SV} and V_{SH} perturbations of $\sim 5\%$ is obtained for an area just southwest of Hawaii. It is notable that in the global model of Montagner and Tanimoto¹¹, the largest anisotropic signal seen at this depth lies in the same area, though it has a smaller absolute amplitude. Below 150 km depth, the difference between V_{SH} and V_{SV} beneath the central Pacific becomes smaller (Fig. 2d), and we do not believe we can resolve a difference between the two at depths greater than 250 km with our current data set.

Two main conclusions result from our analysis. First, Fig. 1 shows that for most of the world, the anisotropy of PREM provides a good average. The only large region where this does not hold true is the Pacific plate. Figure 2a, b shows average V_{SV} and V_{SH} values calculated for the mantle beneath the Pacific plate and for the rest of the world. The average anisotropy that we obtain, with $V_{SH} \sim 2\text{--}4\%$ faster than V_{SV} between the Moho and 200 km, is similar to that obtained in previous detailed studies of the Pacific upper mantle^{13,14,18}. The Pacific plate is large and well sampled by the

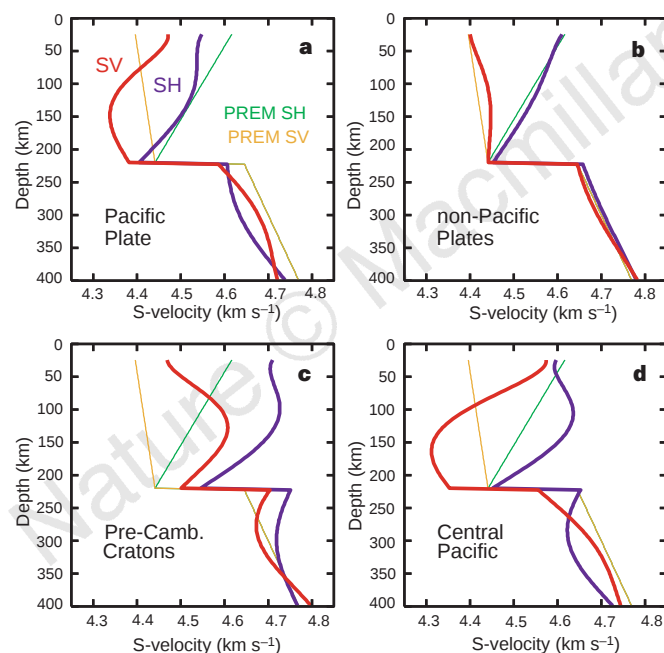


Figure 2 Velocity profiles of the shear-wave velocities V_{SV} and V_{SH} . **a**, The average velocities beneath the entire Pacific plate. The maximum difference between V_{SV} and V_{SH} occurs at around 125 km depth, in contrast to PREM, in which it occurs just below the Moho. Note also that the model shows that the upper mantle beneath the Pacific plate is $\sim 1\%$ slow with respect to PREM down to at least 400 km depth. **b**, The average velocities calculated for all plates except the Pacific. The average structure is close to that of the starting model PREM. **c**, Average velocity profiles calculated for all Precambrian cratons within the Eurasia and North America plates. These regions are sampled very well by our data, and the deviations from the PREM structure are large and well resolved. Note, however, that the difference between V_{SV} and V_{SH} remains very close to that of the starting model PREM. **d**, Average velocity profiles calculated for a cap with 10° radius centred on the anisotropy anomaly identified in this study at $15^\circ\text{N}, 160^\circ\text{W}$. At 130 km depth, V_{SH} is 7% faster than V_{SV} , almost three times the value predicted by PREM. In contrast with PREM, the radial anisotropy in the shallowest mantle is small. Note that the velocity discontinuity at 220 km is part of the starting model PREM, and not a feature that can be resolved in our inversion.

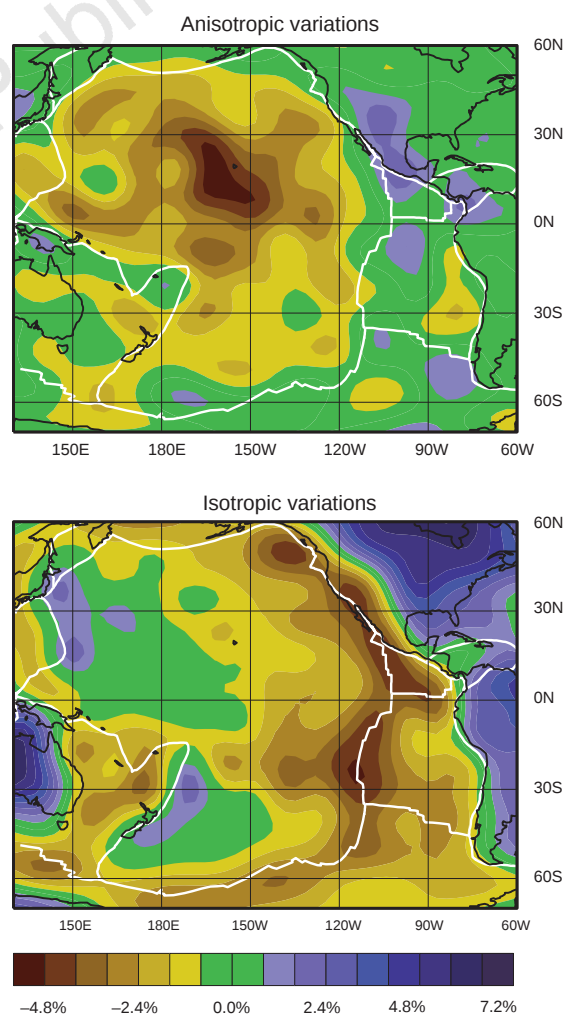


Figure 3 Shear-wave velocity variations beneath the Pacific plate at 150 km depth. The top panel shows the anisotropy $\delta V_{SV} - \delta V_{SH}$ on the same scale as the Voigt-averaged isotropic variation in S-wave velocity (bottom panel) calculated using the approximate expression $\delta V_{\text{Voigt}} = \frac{1}{3}(\delta V_{SH} + 2\delta V_{SV})$. The maps clearly illustrate that anisotropic velocity variations are as large as the isotropic (thermal) variations. The isotropic S-wave variations correlate better with the age of the ocean floor than either the δV_{SV} or δV_{SH} maps in Fig. 1. The largest deviation from this age correlation is very clearly associated with the location of the Pacific Superswell⁶.

data, and we infer that the characteristic anisotropy shown in Fig. 1 is well resolved and unique to the Pacific plate. The North American and Eurasian continents, which are also very well sampled by the data, show very large velocity anomalies with respect to PREM but, surprisingly, very small deviations from the simple anisotropy of PREM (Figs 1 and 2c). The observation that radial anisotropy is required for continental structure agrees with several previous studies^{12,15,20,21}, though it disagrees with the results of some other workers²², in particular those of Montagner and Tanimoto¹¹, who concluded that globally radial anisotropy beneath continents is small.

Our second conclusion is that the geographical variations in radial anisotropy beneath the Pacific plate are large, and do not appear to be correlated with the age of the sea floor (Fig. 3, top map). This large difference between V_{SH} and V_{SV} complicates the interpretation of mapped seismic velocity anomalies²³, as neither V_{SV} nor V_{SH} heterogeneity reflects thermal effects when variations in anisotropy are large. The anisotropic elastic parameters can be combined to form a Voigt average shear modulus¹⁹, which better reflects the isotropic variations in elastic properties—the result of this calculation is shown in Fig. 3. It is striking how well the progression from slow to fast velocities shown in the isotropic average map correlates with the age of the sea floor, with a few notable exceptions: the ‘Pacific Superswell’²³ shows velocities which are as slow as those beneath the EPR, and a small, slow anomaly can be seen to the northwest of Hawaii.

The origin of the Pacific anisotropy anomaly shown in Fig. 3 is unknown. Many mantle rocks and minerals show variable and strong anisotropic elastic effects when examined in the laboratory²⁴; lattice-preferred orientation (LPO) of olivine due to the shearing of mantle materials during the formation and translation of the plate has been suggested as the main mechanism for generating large-scale seismic anisotropy in both the lithosphere^{13,20,25} and the underlying asthenosphere^{11,18,26}. However, the generation of radial anisotropy by LPO depends on an azimuthally random, but preferentially horizontal, alignment of olivine crystals. It is unclear what flow and shearing mechanisms exist to accomplish this beneath the middle of the Pacific plate. One possibility is that, in some areas, the shearing in the asthenosphere is not dominated by the coherent plate motions, but instead by processes with characteristic length scales significantly smaller than the resolving length of our model, ~1,500 km. Small-scale convection²⁷, or a complex pattern of horizontal flow generated by the injection of material into the asthenosphere by mantle plumes, could potentially cause such small-scale variability in the aligning shear motions.

Table 1 Data used in deriving the anisotropic model S20A

Data type	Observations	Sensitivity	Variance reduction
Travel times			
SS-S	5,124	SV	72%
ScS-S	3,543	SV	61%
S-SKS	3,567	SV	55%
SKKS-SKS	2,232	SV	32%
S	26,462	SV	37%
SS	11,417	SV	39%
ScS	4,422	SV	49%
ScSScS	1,230	SV	40%
Long-period waveforms			
Body waves, $T > 45$ s (Z, L)	13,997	SV	28%
Mantle waves, $T > 85$ s (Z, L)	5,523	SV	55%
Mantle waves, $T > 85$ s (T)	2,665	SH	59%
Mantle waves, $T > 135$ s (Z, L)	4,422	SV	40%
Mantle waves, $T > 135$ s (T)	1,710	SH	47%
Mantle waves, $T > 200$ s (Z, L)	4,978	SV	14%
Mantle waves, $T > 200$ s (L)	1,853	SH	23%
Surface wave dispersion			
Rayleigh (R1), 35–150 s	27,948–37,240	SV	68–90%
Love (G1), 35–150 s	15,189–23,947	SH	76–96%
Rayleigh (R1, R2), 150–300 s	18,295–10,256	SV	30–71%
Love (G1, G2), 150–300 s	9,143–3,830	SH	45–73%

See Methods for details of the data shown here.

Although seismic tomography provides a tool for mapping elastic properties of the Earth’s interior, it is increasingly clear that the assumptions driving the interpretation of these maps must continue to be questioned. Composition and fabric may, in some regions of the mantle, be as important as thermal effects in generating velocity anomalies. The positive corollary of this complexity is that the estimation of these additional properties may ultimately be of equal importance for inferring the dynamical processes occurring in the deep Earth. □

Methods

Our model parametrization describes 3D S-wave velocity perturbations with respect to PREM¹⁹. As in earlier studies²⁸, the model is expanded using basis functions:

$$\delta V_S(r, \theta, \phi)/V_0(r) = \sum_{k=0}^K \sum_{l=0}^L \sum_{m=0}^l f_k(r) \times p_l^m(\cos\theta) ({}_kA_l^m \cos m\phi + {}_kB_l^m \sin m\phi) \quad (1)$$

where $V_0(r)$ is the PREM reference velocity, $f_k(r)$ are normalized Chebyshev polynomials, $p_l^m(\cos\theta)$ the associated Legendre polynomials, θ is colatitude, ϕ is longitude, and r is the radius. The upper and lower mantle are described by separate sets of coefficients, and the maximum degree in the radial expansion is $K = 7$ for the upper mantle and $K = 5$ for the lower mantle. In the horizontal direction, the maximum spherical harmonic degree is $L = 20$, for both the upper and lower mantle. (Coefficients describing model S20A are available; see Supplementary Information.)

Table 1 lists the four types of data used in the model inversion. (1) Absolute and differential travel times measured from digital seismograms using a cross-correlation technique²⁸. (2) Complete long-period waveforms in several frequency bands. These data are similar to the data used in the early Harvard tomographic models of the upper mantle²⁹, but the data set has been much expanded. (3) Surface-wave dispersion measurements³⁰ in the period range 35–150 s. (4) Additional long-period (150–300 s) Love and Rayleigh wave dispersion measurements. The data sets of greatest importance for resolving the anisotropy of the upper mantle are the measurements of Love and Rayleigh wave dispersion, data types (3) and (4).

Five elastic parameters are required to describe a transversely isotropic medium³¹. Using the theory developed by Takeuchi and Saito³² and Woodhouse and Dziewonski²⁹, one could attempt an inversion for all five elastic constants, A , C , F , L and N in the notation of Love³¹, but our data have only limited independent sensitivity to the first three parameters. Thus, an inversion focused on 3D variations in $V_{SH} = (N/\rho)^{1/2}$ and $V_{SV} = (L/\rho)^{1/2}$ seemed at this stage most practical and straightforward, where ρ is the density. To simplify the calculations, we used average isotropic sensitivity kernels for δV_S calculated for anisotropic PREM in our inversions. We have performed experiments which show that no significant bias is introduced by this approximation (see below). The sensitivity of Rayleigh waves to independent variations in V_{PH} and V_{PV} is not considered in our analysis, but in the calculation of the sensitivity kernels we assume an implicit variation in P-wave velocity such that $\delta V_P/V_P = 0.55\delta V_S/V_S$. Additional experiments using complete anisotropic kernels show that varying the proportionality constant between 0.0 and 1.0 does not significantly affect our results. (Details of these, and the above-mentioned experiments are available; see Supplementary Information.) Because the sensitivity of Rayleigh waves to P-wave velocity is much smaller than that to S-wave velocity, we believe that unrealistically large independent variations in P-wave velocity would be required to influence our main result.

To constrain the V_{SH} and V_{SV} velocities independently in the upper mantle, we divided the data sets into two groups: one with primary sensitivity to V_{SH} and a second with primary sensitivity to V_{SV} . Travel-time data were included in the V_{SV} data set as all measured phases bottom in the lower mantle and have nearly vertical propagation paths in the upper mantle. Rayleigh-wave dispersion and mantle- and body-wave data observed on the vertical and longitudinal components were also included in the V_{SV} data set. Only fundamental-mode Love wave dispersion and mantle-wave data were included in the V_{SH} data set.

The model coefficients (${}_kA_l^m$, ${}_kB_l^m$) were determined by formulating the inverse problem, which minimizes the least-squares misfit between the observed data and corresponding model predictions. The data were inverted

simultaneously for isotropic velocity perturbations in the lower mantle and independent V_{SV} and V_{SH} variations in the upper mantle. Because the true observational and modelling errors are unknown, the different data sets are combined using empirical weighting coefficients based on previous experiments and our subjective judgement. The inversion was damped towards a minimum model and smooth horizontal perturbations. The perturbations were also damped towards continuity across the 670-km discontinuity. No explicit radial damping was imposed in the inversion.

An important step in imaging the shallow mantle is the application of corrections for crustal structure. We correct the Love and Rayleigh wave dispersion curves by subtracting the predicted effect of the crustal model CRUST-5.1 (ref. 33). Travel-time data are also corrected for this crustal model, as well as for topography and bathymetry.

Azimuthal anisotropy, the dependence of velocity on the direction of wave propagation^{13,18}, is observed to be significant for Rayleigh waves traversing oceanic lithosphere, and could potentially bias our results. We have therefore inverted the Rayleigh wave dispersion data to obtain maps of azimuthal anisotropy, and corrected the data for the azimuthally varying term when inverting for 3D Earth structure. The application of this correction does not, however, materially change the derived 3D models.

The reductions in variance provided by the 3D model S20A with respect to the starting model (PREM) for the various data sets are given in Table 1.

Received 24 July 1997; accepted 6 April 1998.

- Zhang, Y.-S. & Tanimoto, T. Ridges, hotspots and their interaction as observed in seismic velocity maps. *Nature* **355**, 45–49 (1992).
- Su, W.-J., Woodward, R. L. & Dziewonski, A. M. Deep origin of mid-ocean-ridge seismic velocity anomalies. *Nature* **359**, 149–152 (1992).
- McNutt, M. K. & Superswells, Rev. *Geophys. Space Phys.* (in the press).
- Ekström, G. & Dziewonski, A. M. Improved models of upper mantle S velocity structure. (abstr.) *EOS* **76**(46), F421 (1995).
- Su, W.-J., Woodward, R. L. & Dziewonski, A. M. Degree-12 model of shear velocity heterogeneity in the mantle. *J. Geophys. Res.* **99**, 4945–4980 (1994).
- Masters, G., Johnson, S., Laske, G. & Bolton, H. A shear-velocity model of the mantle. *Phil. Trans. R. Soc. Lond. A* **354**, 1385–1411 (1996).
- Li, X.-D. & Romanowicz, B. Global mantle shear velocity model developed using nonlinear asymptotic coupling theory. *J. Geophys. Res.* **101**, 22245–22272 (1996).
- Nataf, H. C., Nakanishi, I. & Anderson, D. L. Anisotropy and shear-velocity heterogeneities in the upper mantle. *Geophys. Res. Lett.* **11**, 109–112 (1984).
- Nataf, H. C., Nakanishi, I. & Anderson, D. L. Measurements of mantle wave velocities and inversion for lateral heterogeneities and anisotropy. *J. Geophys. Res.* **91**, 7261–7307 (1986).
- Montagner, J.-P. & Tanimoto, T. Global anisotropy in the upper mantle inferred from the regionalization of phase velocities. *J. Geophys. Res.* **95**, 4797–4819 (1990).
- Montagner, J.-P. & Tanimoto, T. Global upper mantle tomography of seismic velocities and anisotropies. *J. Geophys. Res.* **96**, 20337–20351 (1991).
- McEvilly, T. V. Central US crust-upper mantle structure from Love and Rayleigh wave phase velocity inversion. *Bull. Seismol. Soc. Am.* **54**, 1997–2015 (1964).
- Forsyth, D. W. The early tectonic evolution and anisotropy of the oceanic upper mantle. *Geophys. J. R. Astron. Soc.* **43**, 103–162 (1975).
- Mitchell, B. J. & Yu, G.-K. Surface wave dispersion, regionalized velocity models, and anisotropy of the Pacific crust and upper mantle. *Geophys. J. R. Astron. Soc.* **63**, 497–514 (1980).
- Gee, L. S. & Jordan, T. H. Polarization anisotropy and fine-scale structure of the Eurasian upper mantle. *Geophys. Res. Lett.* **15**, 824–827 (1988).
- Gaherty, J. B., Jordan, T. H. & Gee, L. S. Seismic structure of the upper mantle in a central Pacific corridor. *J. Geophys. Res.* **101**, 22291–22309 (1996).
- Nishimura, C. E. & Forsyth, D. W. Rayleigh wave phase velocities in the Pacific with implications for azimuthal anisotropy and lateral heterogeneities. *Geophys. J. R. Astron. Soc.* **94**, 479–501 (1988).
- Nishimura, C. E. & Forsyth, D. W. The anisotropic structure of the upper mantle in the Pacific. *Geophys. J. R. Astron. Soc.* **96**, 203–229 (1989).
- Dziewonski, A. M. & Anderson, D. L. Preliminary reference Earth model. *Phys. Earth Planet. Inter.* **25**, 297–356 (1981).
- Léveque, J. J. & Cara, M. Long-period Love wave overtone data in North America and the Pacific Ocean: new evidence for upper mantle anisotropy. *Phys. Earth Planet. Inter.* **33**, 164–179 (1983).
- Gaherty, J. B., Kato, M. & Jordan, T. H. Seismological structure of the upper mantle: A regional comparison of seismic layering. *Phys. Earth Planet. Inter.* (submitted).
- Mitchell, B. J. On the inversion of Love and Rayleigh wave dispersion and implications for earth structure and anisotropy. *Geophys. J. R. Astron. Soc.* **76**, 233–241 (1984).
- Regan, J. & Anderson, D. L. Anisotropic models of the upper mantle. *Phys. Earth Planet. Inter.* **35**, 227–263 (1984).
- Birch, F. The velocity of compressional waves in rocks to 10 kilobars. *J. Geophys. Res.* **65**, 1083–1102 (1960).
- Nicolas, A. & Christensen, N. I. In *Composition, Structure, and Dynamics of Lithosphere-Asthenosphere System* (eds Fuchs, K. & Froidevaux, C.) 111–123 (Geodyn. Ser. Vol. 16, Am. Geophys. Union, Washington DC, 1987).
- Ribe, N. M. Seismic anisotropy and mantle flow. *J. Geophys. Res.* **94**, 4213–4223 (1989).
- Parsons, B. & McKenzie, D. Mantle convection and the thermal structure of the plates. *J. Geophys. Res.* **83**, 4485–4496 (1978).
- Dziewonski, A. M., Ekström, G. & Liu, X.-F. In *Monitoring a Comprehensive Test Ban Treaty* (eds Husebye, E. S. & Dainty, A. M.) 521–550 (Kluwer Academic, Dordrecht, 1996).
- Woodhouse, J. H. & Dziewonski, A. M. Mapping the upper mantle: Three dimensional modeling of Earth structure by inversion of seismic waveforms. *J. Geophys. Res.* **89**, 5953–5986 (1984).
- Ekström, G., Tromp, J. & Larson, E. W. F. Measurements and global models of surface wave propagation. *J. Geophys. Res.* **102**, 8317–8157 (1997).
- Love, A. E. H. *A Treatise on the Theory of Elasticity* 4th edn (Cambridge Univ., 1927).
- Takeuchi, H. & Saito, M. Seismic surface waves. *Methods Comput. Phys.* **11**, 217–295 (1972).

33. Mooney, W. D., Laske, G. & Masters, G. CRUST-5.1: A global crustal model at $5^\circ \times 5^\circ$. *J. Geophys. Res.* **103**, 727–747 (1998).

Supplementary information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. The data used in this work were obtained from the IRIS, GDSN, IDA, GEOSCOPE, MEDNET, CDSN and GTSN seismograph networks. We thank D. Forsyth, J. Gaherty, J. Phipps Morgan, G. Smith and C. Wolfe for discussions, and X.-F. Liu, S. Sianissian and W.-J. Su for help with collecting and preparing the data sets. This work was supported by the US NSF and the US Air Force Office for Scientific Research.

Correspondence and requests for materials should be addressed to G.E. (e-mail: ekstrom@seismology.harvard.edu).

Devonian terrestrial arthropods from Gondwana

Gregory D. Edgecombe

Centre for Evolutionary Research, Australian Museum, 6 College Street, Sydney South, New South Wales 2000, Australia

The origin of the most diverse terrestrial animal group, Atelocerata (myriapods and hexapods), is obscured by an incomplete fossil record¹. Early (Silurian and Devonian) body fossils of terrestrial arthropods have been found only in Laurussia, with key sites in Britain and eastern North America^{2–5}. Although trace fossil assemblages indicate the presence of various arthropods on land in Australia in the Silurian Period⁶, definite terrestrial arthropods have not been discovered in mid-Palaeozoic stages of the southern continents. Here I describe the first atelocerates from the Devonian stages of Gondwana; these are perhaps the earliest known remains of Australian land animals. The fossils comprise two closely related myriapod species of the genus *Maldybulakia*, first identified from Kazakhstan^{7,8}. They add substantially to our knowledge of the anatomy of this problematic arthropod, and illustrate the widespread distribution of parts of the Devonian terrestrial fauna. A clade including *Maldybulakia* is distinct within the Myriapoda at a high taxonomic level. The existence of *Maldybulakia* and the extinct classes Arthropleuridea and Kampekarida⁹, with centipedes and millipedes, indicates the high class-level diversity of myriapods in the Devonian.

The earlier of two new Devonian species of *Maldybulakia* from New South Wales, Australia, is represented by abundant moulds of disarticulated tergites and pleurotergites in sandstones of the Sugarloaf Creek Formation, in the Taemas–Wee Jasper area. These strata are of Early Devonian (latest Lochkovian to earliest Pragian) age, and have a fluvial origin¹⁰. The second Australian species is found in lacustrine mudstones of the Boyd Volcanic Complex near Eden; these mudstones have been assigned a Late Givetian or Early Frasnian age (ref. 11, and G. C. Young, personal communication). At both sites they are the only animal macrofossils, but the Eden locality has an abundant terrestrial flora. *Maldybulakia* is known from lacustrine facies in the Lower Devonian (Pragian/Emsian) stages of central Kazakhstan⁷.

Articulated specimens of *Maldybulakia* (Fig. 1a, b) show the trunk to be composed of two tagmata. The anterior tagma includes one or possibly two trapezoidal tergites. This tergite type is bisected medially by a sinuous, transverse furrow or stricture, and has a sloping anterior section. Although only one of these tergites is found in several articulated specimens of one species, another of the same size is displaced in one specimen, and the considerable variation in this tergite in the other species (Fig. 2a) is suggestive of more than one tergite in this tagma. The posterior trunk tagma is composed of pleurotergites with triangular lateral lobes set off from the posteromedian part of the tergum by deep, diagonal furrows. The anterior part of the tergites is lenoid in outline (Fig. 1e) and is completely overlapped by the preceding pleurotergite in articulation. I accept earlier description of the anterior and posterior parts of these

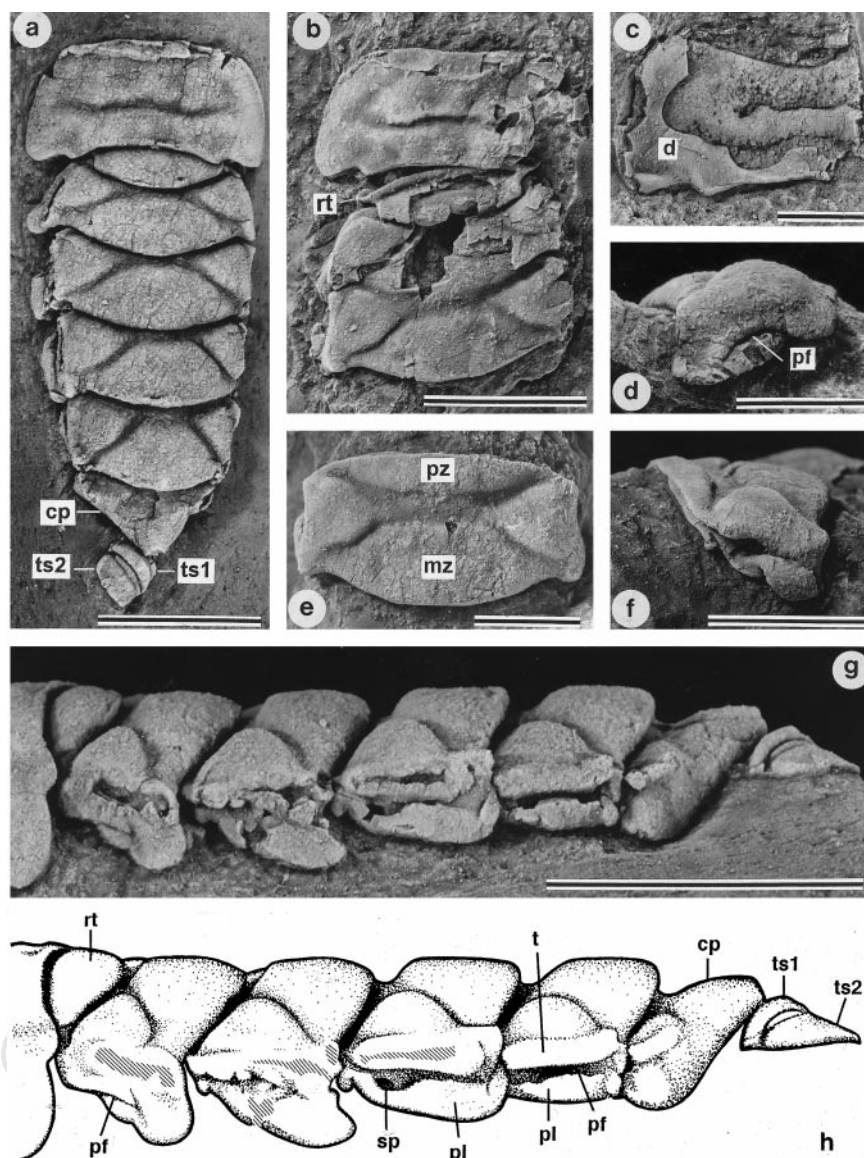


Figure 1 *Maladybulakia* new species 1 from the Middle or Upper Devonian (Late Givetian or Early Frasnian) Epochs at Saltwater Creek Road, 12 km south-south-east of Eden, New South Wales, Australia. All scale bars represent 1 cm. Specimens are lodged in the Palaeontology type collections, Australian Museum. **a**, Dorsal view of trunk, specimen AM F:102533, showing one tergite of the anterior tagma, four pleurotergites of the posterior tagma, a caudal pleurotergite (cp), and a disarticulated pair of telson sclerites (ts1 and ts2). **b**, Dorsal view of a tergite of the anterior trunk tagma and the first two pleurotergites of the posterior trunk tagma, AM F:102534. A ringlike tergite (rt) provides the articulation between the

two tagmata. **c**, Ventral view of a tergite of the anterior trunk tagma, AM F:102550, showing doubleure (d). **d**, Posterolateral view of the first pleurotergite of the posterior trunk tagma, AM F:102552. Note the pleural furrow (pf). **e**, **f**, Dorsal (**e**) and lateral (**f**) views of the second pleurotergite of the posterior trunk tagma, AM F:102555. Prozonite (pz) and metazonite (mz) are indicated. **g**, Lateral view of the trunk, showing pleurites fused to metazonites of diplotergites. **h**, Interpretive drawing of **g**. Breaks in the exoskeletal surface are indicated by cross-hatching. Pl, pleurite; sp, spiracle; t, lateral edge of tergum.

pleurotergites as prozonites and metazonites of diplosegments⁷, as also applies to the divided anterior tergite (Fig. 1c). Four of the lobate pleurotergites are present in articulated specimens of one species (Fig. 1a). The number of pleurotergites is unknown in the other species, members of which have long paratergal spines (Fig. 2b, c). The prozonite of the anteriormost lobate pleurotergite is overlapped by a ringlike tergite that has socket-like lateral projections (Fig. 1b). These projections lie against bulges on the posterior margin of the trapezoidal tergite and act as articulations between the two tagmata.

A major topographic break from the tergum suggests that lateral sclerotizations (Fig. 1d, f–h) are pleural rather than paratergal. The pleurites are fused to the tergum on the metazonites of the posterior tagma, and are separated from the folded edge of the tergum by a

pleural furrow. The anterior part of the pleural furrow of the third pleurotergite is expanded, with an ovate, cuticle-lined canal projecting inwards (Fig. 1g, h). This canal is identified as a spiracle, apparently the only such spiracle on the trunk.

Following the four lobate pleurotergites is a caudal pleurotergite with a conical or spinose posteromedian projection. A pair of small sclerites, the anterior one being ringlike and the posterior broadly quadrate (Fig. 1a, g), are evidently the terminal part of the exoskeleton, and regarded as the telson.

The tergum is strongly mineralized, with the cuticle being perforated by closely spaced canals that open to large pores on the ventral surface. The absence of sternites and legs amidst hundreds of tergites/pleurotergites for both Australian species indicates that the sternum and legs were unmineralized. A head has not been found at

either site; the most anterior known tergites conform more closely to trunk diploptergites than to a cephalic shield. The head is presumed to have had a membranous attachment to the trunk and undergone a different taphonomic history.

Exoskeletal features of *Maldybulakia* best support an assignment to Atelocerata and in particular to the myriapods, a group variably regarded as monophyletic^{12,13} or paraphyletic¹⁴. Monophyly of Atelocerata is well supported by morphological characters^{14,15}; conflicting molecular phylogenies^{16,17} may be due to alignment or sampling defects¹⁸, and are not robust to a combined data approach or a general range of analytical parameters¹⁹. Most of the anatomical characters that define Atelocerata (or those alleged to define Myriapoda) are not preserved in *Maldybulakia*. The interpretation of the tubular opening in the pleural region as a spiracle is, however, evidence for a tracheal respiratory system as in myriapods and hexapods, and implies that *Maldybulakia* was terrestrial. Pleurites are present in *Maldybulakia*, and are characteristic of (perhaps unique to²⁰) atelocerates; their fusion to form pleurotergites is shared by diplopods.

The apparent diplosegmental construction of the trunk in *Maldybulakia* indicates possible closest affinities to Dignatha, and in particular the Diplopoda, among extant taxa. A strong tergal exoskeleton is shared with diplopods and the Silurian/Devonian

Kampecarida, which appear to be diplopod-allied on the basis of a sternal attachment of the legs and presence of anterior haplotergites^{9,21}. A closer relationship between kampecarids and *Maldybulakia* is suggested by large cuticular canals and pores, and an unmineralized sternum²¹.

The most plausible rival to myriapod affinities for *Maldybulakia* would be a crustacean assignment, notably to malacostracans such as isopods. This follows more from the broad morphological spectrum exhibited by crustacean fossils in freshwater facies than from any unique crustacean characters seen in *Maldybulakia*. Affinities to many crustacean taxa are weakened by the lack of a carapace at any site, despite different taphonomic histories. Articulated specimens that are complete posteriorly (Fig. 1a) possess too few tergites to accommodate the postcephalic segmentation of malacostracans (eight pereonites and seven pleonites).

We recognize Arthropleuridea (Silurian/Permian^{22,23}) and Kampecarida as extinct classes of Myriapoda⁹, with the same taxonomic rank as the extant Diplopoda, Pauropoda, Chilopoda, and Symphyla; thus it appears that class-level diversity of the Myriapoda peaked in the Palaeozoic Era. A clade including *Maldybulakia* can be distinguished from other myriapod classes by the few trunk segments, marked trunk tagmosis, lobation of the tergites, and a single pair of spiracles in the pleural furrow. All classes of Myriapoda have fossil records or ghost lineages in the Devonian, drawing the acme of myriapod body-plan diversity into this period.

On the basis of trace fossil evidence⁶, terrestrial arthropods have a Gondwanan history that pre-dates the Early Devonian occurrence of *Maldybulakia*. The only arthropod remains known from the Gondwanan trace fossil assemblages, euthycarcinoids²⁴, have been associated with subaerial trackways⁶. However, the limbs of better preserved euthycarcinoids from other locations have non-plantigrade structure, lack pretarsal claws, and bear a strong fringe of setae, indicative of wholly aquatic habits^{4,21}.

The disjointed biogeographic distribution of *Maldybulakia* predicts as yet unsampled occurrences in terranes between Kazakhstan and Australia. Lower Devonian palaeocontinental reconstructions that show a continuous land bridge from Kazakhstan to Australia²⁵, linked by North China, agree better with the distribution of *Maldybulakia* than do reconstructions with significant oceanic separation of Kazakhstan and Australia²⁶. Earliest Devonian (Lochkovian) reconstructions with a narrow sea separating Kazakhstan and North China²⁵ indicate that the dispersal of *Maldybulakia* is likely to have occurred during the Pragian or Emsian, and species might be expected in appropriate facies in the Lower Devonian of North China. The distribution of *Maldybulakia* indicates that parts of the Devonian terrestrial fauna may have been widely distributed in and around those terranes derived from Gondwana. □

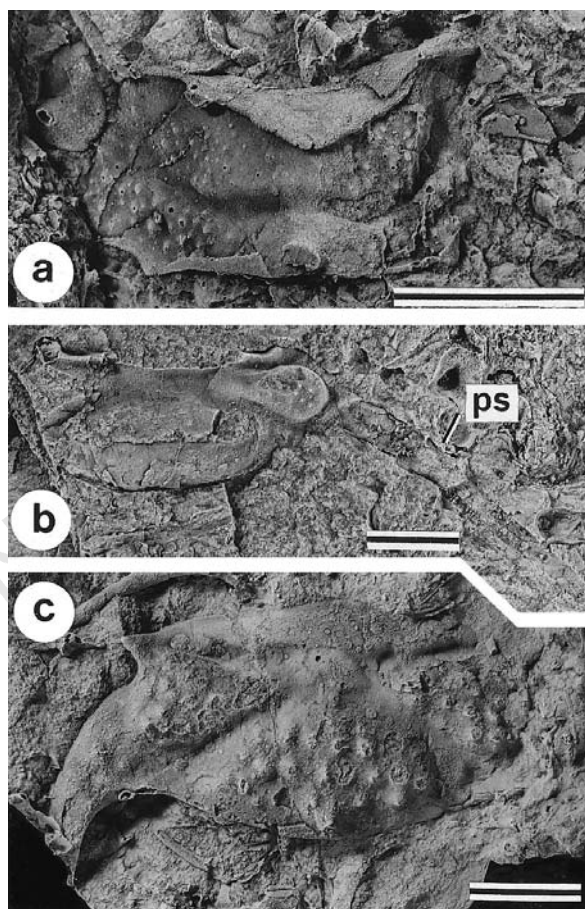


Figure 2 *Maldybulakia* new species 2 from the Lower Devonian (latest Lochkovian or earliest Pragian) Epoch at Three Oaks, near Taemas, New South Wales, Australia. All scale bars represent 1 cm. **a**, Dorsal view of a tergite from the anterior trunk tagma, specimen AM F:102563; latex cast from external mould. Note median spine on metazonite. **b**, Dorsal view of the first pleurotergite from the posterior trunk tagma, AM F:102570b; internal mould. The long paratergal spines (ps) and coarse tuberculation are unique to this species. **c**, Dorsal view of a pleurotergite from the middle part of the posterior trunk tagma, AM F:102565a; latex cast from external mould.

Received 4 December 1997; accepted 26 March 1998.

1. Rolfe, W. D. I. Early terrestrial arthropods: a fragmentary record. *Phil. Trans. R. Soc. Lond. B* **309**, 207–218 (1985).
2. Almond, J. E. The Silurian–Devonian fossil record of the Myriapoda. *Phil. Trans. R. Soc. Lond. B* **309**, 227–237 (1985).
3. Jeram, A. J., Selden, P. A. & Edwards, D. Land animals in the Silurian: arachnids and myriapods from Shropshire, England. *Science* **250**, 658–661 (1990).
4. Shear, W. A. in *Arthropod Paleobiology. Short Courses in Paleontology* 3 (ed. Mikulic, D. G.) 197–213 (Paleontol. Soc., Knoxville, 1990).
5. Shear, W. A., Gensel, P. G. & Jeram, A. J. Fossils of large terrestrial arthropods from the Lower Devonian of Canada. *Nature* **384**, 555–557 (1996).
6. Trevis, N. H. & McNamara, K. J. Arthropods invade the land: trace fossils and palaeoenvironments of the Tumblagooda Sandstone (?late Silurian) of Kalbarri, Western Australia. *Trans. R. Soc. Edinb. Earth Sci.* **85**, 177–210 (1995).
7. Tesakov, A. S. & Alekseev, A. S. Myriapod-like arthropods from the Lower Devonian of central Kazakhstan. *Paleontol. J.* **26**, 18–23 (1992).
8. Tesakov, A. S. & Alekseev, A. S. *Maldybulakia*—new name for *Lophodesmus* Tesakov & Alekseev, 1992 (Arthropoda). *Paleontol. J.* **32**, 49 (1998).
9. Shear, W. A. in *Arthropod Relationships* (eds Fortey, R. A. & Thomas, R. H.) 211–219 (Chapman & Hall, London, 1997).
10. Owen, M. & Wyborn, D. Geology and geochemistry of the Tantangara and Brindabella area. *Bull. Bur. Miner. Resour. Geol. Geophys. Aust.* **204** (1979).
11. Fergusson, C. L. et al. The Upper Devonian Boyd Volcanic Complex, Eden, New South Wales. *J. Geol. Soc. Aust.* **26**, 87–105 (1979).
12. Regier, J. C. & Shultz, J. W. Molecular phylogeny of the major arthropod groups indicates polyphyly of crustaceans and a new hypothesis for the origin of hexapods. *Mol. Biol. Evol.* **14**, 902–913 (1997).

13. Zrzavý, J., Hýpša, V. & Vlášková, M. in *Arthropod Relationships* (eds Fortey, R. A. & Thomas, R. H.) 97–107 (Chapman & Hall, London, 1997).
14. Kraus, O. in *Arthropod Relationships* (eds Fortey, R. A. & Thomas, R. H.) 295–303 (Chapman & Hall, London, 1997).
15. Wheeler, W. C., Cartwright, P. & Hayashi, C. Y. Arthropod phylogeny: a combined approach. *Cladistics* **9**, 1–39 (1993).
16. Ballard, J. W. et al. Evidence from 12S ribosomal RNA sequences that onychophorans are modified arthropods. *Science* **258**, 1345–1348 (1992).
17. Friedrich, M. & Tautz, D. Ribosomal DNA phylogeny of the major extant arthropod classes and the evolution of myriapods. *Nature* **376**, 165–167 (1995).
18. Wägele, J. W. & Stanjek, G. Arthropod phylogeny inferred from partial 12S rRNA revisited: monophyly of the Tracheata depends on sequence alignment. *J. Zool. Syst. Evol. Res.* **33**, 75–80 (1995).
19. Wheeler, W. C. in *Arthropod Fossils and Phylogeny* (ed. Edgecombe, G. D.) 9–32 (Columbia Univ. Press, New York, 1998).
20. Mantou, S. M. in *Myriapod Biology* (ed. Camatini, M.) 317–343 (Academic, London, 1979).
21. Almond, J. E. *Studies on Palaeozoic Arthropoda*. Thesis, Univ. Cambridge (1986).
22. Shear, W. A. & Selden, P. *Eoarthropleura* (Arthropoda, Arthropleurida) from the Silurian of Britain and the Devonian of North America. *Neues. Jb. Geol. Palaeontol. Abh.* **196**, 347–375 (1995).
23. Schneider, J. Environment, biotas and taphonomy of the Lower Permian lacustrine Neiderhäslich limestone, Döhlen Basin, Germany. *Trans. R. Soc. Edinb. Earth Sci.* **84**, 453–464 (1994).
24. McNamara, K. J. & Trewin, N. H. A eurycarinoid arthropod from the Silurian of Western Australia. *Palaeontology* **36**, 319–335 (1993).
25. Scotese, C. R. & McKerrow, W. S. in *Palaeozoic Paleogeography and Biogeography* (eds McKerrow, W. S. & Scotese, C. R.) 1–21 (Geol. Soc. Lond., 1990).
26. Li, Z.-X., Powell, C. M. & Trench, A. in *Palaeozoic Vertebrate Biostratigraphy and Biogeography* (ed. Long, J. A.) 25–53 (Belhaven, London, 1993).

Acknowledgements. I thank K. S. W. Campbell and A. Ritchie for showing me the fossil sites; Z. Johanson, B. Loomes, Y.-y. Zhen and volunteers organized by M. Yeung for assistance in collecting fossils; and D. E. G. Briggs, J. A. Dunlop, W. A. Shear, and G. D. F. Wilson for comments. NSW State Forests permitted collecting. A grant from Australian Geographic supported fieldwork.

Correspondence and requests for materials should be addressed to the author (e-mail: greged@ams.g.austmus.gov.au).

Development of cooperative relationships through increasing investment

Gilbert Roberts* & Thomas N. Sherratt†

* Evolution and Behaviour Research Group, Department of Psychology, University of Newcastle upon Tyne, Newcastle upon Tyne NE1 7RU, UK

† Department of Biological Sciences, University of Durham, South Road, Durham DH1 3LE, UK

Reciprocal altruism¹ can become established among selfish, unrelated individuals if they use responsive strategies such as 'tit-for-tat'^{2–4}. This result raises the fundamental question: how altruistic should one be? The problem is difficult to solve using current 'prisoner's dilemma' based models because they allow only the discrete choice of cooperating or defecting. In reality, however, cooperation is rarely all-or-nothing. Furthermore, if cooperative investment is variable, a new and more subtle kind of cheating becomes possible: individuals may invest slightly less than their partner. A concern is that this 'short-changing' will erode cooperative ventures. Here we show that cooperation can thrive despite variable investment through the new strategy of 'raise-the-stakes'. This strategy offers a small amount on first meeting and then, if matched, raises its investment, something that no strategy in the discrete model can do. We show that such behaviour can readily invade a population of non-altruists and cannot be effectively exploited. The practice of 'testing the water' rather than making sudden cooperative 'leaps of faith' powerfully reinforces the stability and effectiveness of reciprocity.

How cooperative should one be? Using current theory², this question can be answered only in terms of the number of cooperative acts performed. But consider a typical reciprocal interaction among guillemots (*Uria aalge*) preening each other. Allopreening bouts vary from under a second to over a minute (G.R., M. P. Harris and S. Wanless, unpublished observations). Conventionally, both extremes are considered as cooperative events but cooperation is not an 'all-or-nothing' behaviour among these allopreeners, or indeed in most instances of apparent reciprocity^{5–7}. Once we allow variable investment, the possibility of a new and more

subtle kind of cheating arises because individuals may invest a little less than their partner¹. This is a potentially serious threat to cooperation, as such short-changing could gradually erode cooperation. Here we show for the first time, to our knowledge, that relationships involving increasing investment are robust in the face of cheats and subtle cheats.

Earlier work recognizing that cooperation is not discrete has been limited to interpolating between possible outcomes in the prisoner's dilemma (PD) payoff matrix⁸, which in itself represents the summed consequence of the behaviours of two interacting individuals. By decomposing the PD payoff matrix, we can focus instead on each individual's behaviour separately. Although, in common usage, 'defection' often implies doing something positively harmful to another party, in reciprocal altruism defection generally corresponds simply to doing nothing (for example, not grooming a partner⁹). Here, therefore, we consider only the costs and benefits involved in altruistic acts. We introduce a scale of investment in altruism, with each act resulting in a variable fitness cost to the altruist of u units. Thus, an investment of zero corresponds to defection in the discrete model. We assume that the more an altruist invests, the more a recipient benefits and that these benefits are given by ku . For reciprocal altruism to result in a net benefit to both, we require $k > 1$. Over a symmetrical course of interactions, the payoffs sum to fit the defining inequalities of the PD: whereas reciprocity pays well, exploitation pays better; and whereas mutual failure to cooperate pays poorly, being exploited pays worst of all.

First, we asked whether cooperation can develop in the presence of cheats and subtle cheats. As in the conventional discrete PD, we consider simple strategies that take into account a partner's previous behaviour, but in our continuous model these strategies can respond quantitatively. As the potential range of strategies becomes enormous, we limit this range by considering strategies representative of five possible responses: never investing, matching a partner's investment, undercutting a partner, escalating investment, and cheating probabilistically. Our representative strategies (Table 1) are: 'non-altruism' (NA), which never invests but will accept altruism; 'give-as-good-as-you-get' (GGG), which matches what its partner last invested; 'short-changer' (SC), which behaves similarly to GGG but invests one less than the partner last invested; 'raise the stakes' (RTS), which escalates its investment with partners that have matched or bettered its last move; 'occasional short-changer' (OSC), which behaves as RTS except that, with a certain probability, it invests one less than pure RTS would; and 'occasional cheat' (OC), which behaves as RTS but, with a certain probability, fails to invest at all. We confine consideration initially to those altruistic strategies that, like tit-for-tat, are 'nice'², investing one unit in altruism when they have no experience of a partner.

The success of the RTS strategy is illustrated through simulation (see Methods). Starting with an even mix of strategies, RTS spreads to fixation (Fig. 1a). It also invades an environment of NA seeded with the other strategies (Fig. 1b), and resists invasion when established (Fig. 1c). As in the discrete model², the number of opportunities for altruism limited the success of altruistic strategies. Nevertheless, beginning with two of each altruistic strategy and population size (N) = 60, RTS spread to fixation in 100% of simulations by generations 13, 10, 9 and 9 for number of rounds (R) = 5, 10, 20 and 100, respectively. These results are also robust to changes in N . With R = 20 and beginning with a population of NA seeded with two of each altruistic strategy, RTS spread to fixation in 100% of simulations by generations 8, 8, 9 and 10 for N = 12, 30, 60 and 100, respectively.

Why does RTS do so well? Like tit-for-tat in the discrete model, RTS is resistant to exploitation by other strategies while making the most of cooperative opportunities. RTS can accrue large scores when playing against itself, scores that are not consistently obtainable by non-altruistic or short-changing strategies (Fig. 2). The passivity of GGG means that although it can do well against RTS, it

does poorly against itself. SC cannot exploit RTS because RTS will not escalate its investment unless matched. The occasional cheating of OSC means that it can exploit RTS to some extent, but it does less well against itself.

For ease of interpretation, we included only representative OC and OSC strategies that cheated on 50% of occasions in the above simulations. To determine whether lower levels of cheating are more successful, we considered an environment in which all individuals played an OC strategy with cheating probabilities ranging from 1 (equivalent to NA), through 0.5, 0.25, 0.1 and 0.05, to 0 (equivalent to pure RTS). Beginning with equal numbers of each strategy and

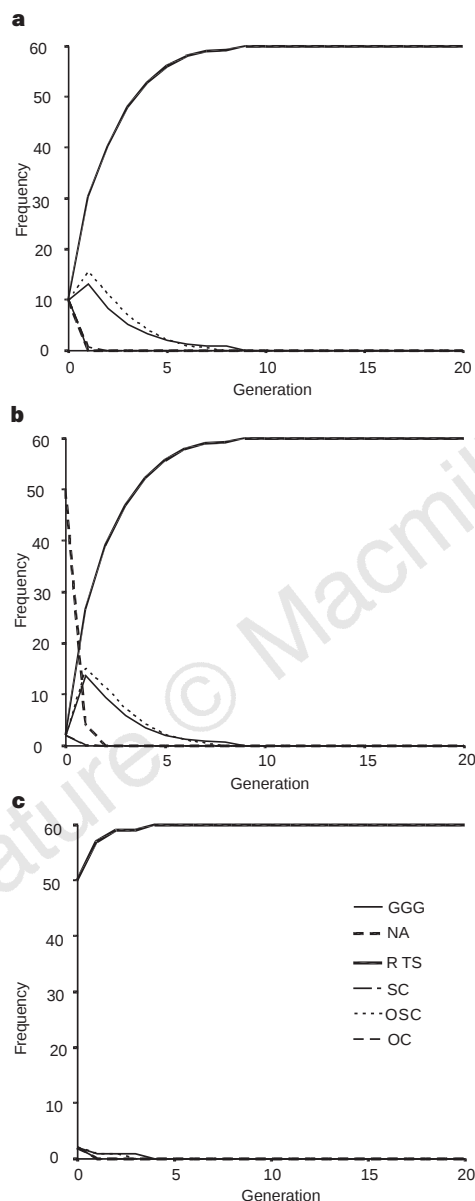


Figure 1 Here we illustrate how a strategy of increasing cooperative investment fares in competition with less altruistic strategies. In each set of simulations, $N = 60$, $R = 20$, $k = 2$ and strategy frequencies in each generation are means over 100 simulations. See Table 1 for strategy definitions. Each strategy except NA had an initial investment of $a = 1$ unit. RTS also had an increment of $b = 1$. Strategies OSC and OC played as $RTS_{1,1}$ but with a probability of cheating of 0.5. **a–c**, The success of the RTS strategy over 20 generations, starting from different initial situations. **a**, All strategies were initially in equal numbers in this case. **b**, RTS invades: the initial population comprised two of each altruistic strategy, the remainder being NA. **c**, RTS resists invasion: the initial population was largely RTS and included two of each other strategy.

with $N = 60$ and $R = 20$, pure RTS emerged by generation 8 in 100% of simulations. Using a corresponding approach, pure RTS emerged from a mix of OSC strategists by generation 46 in 100% of simulations. Thus, the stable probability of cheating is close to zero.

To what extent is the success of RTS due to it escalating such that it can obtain many times its initial returns? Although investment in an activity such as grooming might escalate by a factor of 10 or even 100, there are likely to be limits caused by the escalating costs of altruism and the diminishing returns to the recipient of further altruism. We therefore repeated the set of simulations shown in Fig. 1b but with imposed limits on investment. After 50 generations, and with investment limited to 1, 2, 3, 5 and 10 units, RTS spread to 47, 93, 97, 100 and 100% of the population, respectively. Thus, RTS still tended to invade; however, with a limit of ≤ 3 units RTS came into equilibrium with GGS, which is similar in strategy to RTS when escalating investment is inhibited.

As in the discrete model, a single altruist cannot invade a non-altruistic population. However, we can determine the conditions under which $RTS_{a,b}$ (where a represents the initial investment and b indicates the increment by which the investment escalates) can spread in a population comprising x $RTS_{a,b}$ and y NA strategists when $k > 1$. If all players interact, then $RTS_{a,b}$ will thrive provided that $x > (\phi + ya)/(\phi - ka)$, where $\phi = R(k - 1)(2bR + 2a - b)$. As a typical numerical example, as few as 2 $RTS_{a,b}$ individuals can invade a population of 100 NA individuals when $k = 1.5$, $R = 10$, $a = 1$ and $b = 1$. Thus, the more NA individuals there are the more $RTS_{a,b}$ individuals are required before $RTS_{a,b}$ can invade, but as R increases, the limiting value of x required before $RTS_{a,b}$ can spread is $x > 1$. The signs of the partial derivatives with respect to a (consistently positive) and b (consistently negative) show that fewer $RTS_{a,b}$ individuals are needed to invade a population of NA individuals if the $RTS_{a,b}$ strategists give as little as they can on first meeting (low a) and rapidly escalate if their altruism is reciprocated (high b).

The conditions required for RTS to resist invasion by rare mutants can similarly be determined analytically. Using standard evolutionarily stable strategy (ESS) criteria¹⁰, we can show that an infinitely large population of $RTS_{1,1}$ individuals is stable with respect to invasion by NA individuals when NA is initially rare, provided that $R(k - 1)(2bR + 2a - b) > ka$ (for example, where $k \geq 1.2$ and $R \geq 2$). We can also show that $RTS_{2,1}$ is stable to SC where $k > 1.2$ and $R \geq 2$. We can also ask whether $RTS_{a,b}$ is resistant to invasion by initially more generous forms of RTS (defined by $RTS_{g,b}$ where $g > a$, but $g < (a + 2b)$). As a less generous strategy can potentially gain by subtly cheating a more generous one, $RTS_{a,b}$ will be an ESS if the rare invading variant is too generous, specifically when $(g - a) > b(R + k)/R$.

Is an escalatory strategy like RTS better than one that begins at some intermediate level? To investigate this, we introduced three 'averaging' strategies that, when playing against themselves over R rounds, invest the same average amount r as $RTS_{1,1}$ plays against itself over the same R . The averaging strategies were 'all-or-nothing' (AON), 'anything-will-do' (AWD) and a variant of GGS. They play as defined in Table 1 with $a = r$. As shown in Fig. 3a, in an environment containing a mix of strategies the outcome can be an equilibrium because the cooperative strategies are defined such that they do equally well against themselves and are not easily exploited. However, RTS does particularly well in invading an environment dominated by non-altruistic strategies (Fig. 3b). This is consistent with our earlier result that invasion of NA by RTS is facilitated by lower a . Once established, RTS again exhibits high resistance to exploitation (Fig. 3c), an observation that can be readily confirmed analytically (for instance, $RTS_{1,1}$ is guaranteed to be an ESS against GGS_g (where $g > 1$) if $g > (4k^2 + 4k - 7)/[4(k - 1)]$).

Do animals behave as our model predicts? Some evidence for our prediction that altruists should respond quantitatively to their partners comes from allogrooming in impala (*Aepyceros melam-*

Table 1 Representative strategies

Strategy	Full name	Initial investment	Investment undercut	Response matched	Response bettered
NA	Non-altruism	0	0	0	0
GGG _a	Give-as-good-as-you-get	<i>a</i>	<i>m</i>	<i>m</i>	<i>m</i>
SC	Short-changer	1	<i>m</i> - 1	<i>m</i> - 1	<i>m</i> - 1
RTS _{a,b}	Raise-the-stakes _{a,b}	<i>a</i>	<i>m</i>	<i>p</i> + <i>b</i>	<i>p</i> + 2 <i>b</i>
OSC	Occasional short-changer	RTS or RTS - 1	RTS or RTS - 1	RTS or RTS - 1	RTS or RTS - 1
OC	Occasional cheat	RTS or 0	RTS or 0	RTS or 0	RTS or 0
AWD _a	Anything-will-do	<i>a</i>	<i>a</i>	<i>a</i>	<i>a</i>
AON _a	All-or-nothing	<i>a</i>	0	<i>a</i>	<i>a</i>

Each strategy is defined by its initial investment, that is, the investment made on the leader's first move with a new player, and by how it responds to a partner that has undercut, matched or bettered either its own previous move, or, if none has been made, its own initial generosity. The rules shown are subject to the constraints that all strategies invest zero in response to zero, and that no strategy can invest less than zero. *m* indicates that the agent matches the partner's last investment; *p* indicates that the agent plays as on its own previous move. GGG, RTS, AWD and AON are defined by a parameter *a* indicating their initial investment. RTS has an additional parameter, *b*, indicating the increment by which the investment *a* escalates. OSC and OC are probabilistic strategies which play like RTS₁₁ except in some specified percentage of occasions. References to a strategy without subscripts are to the *a* = 1, *b* = 1 case.

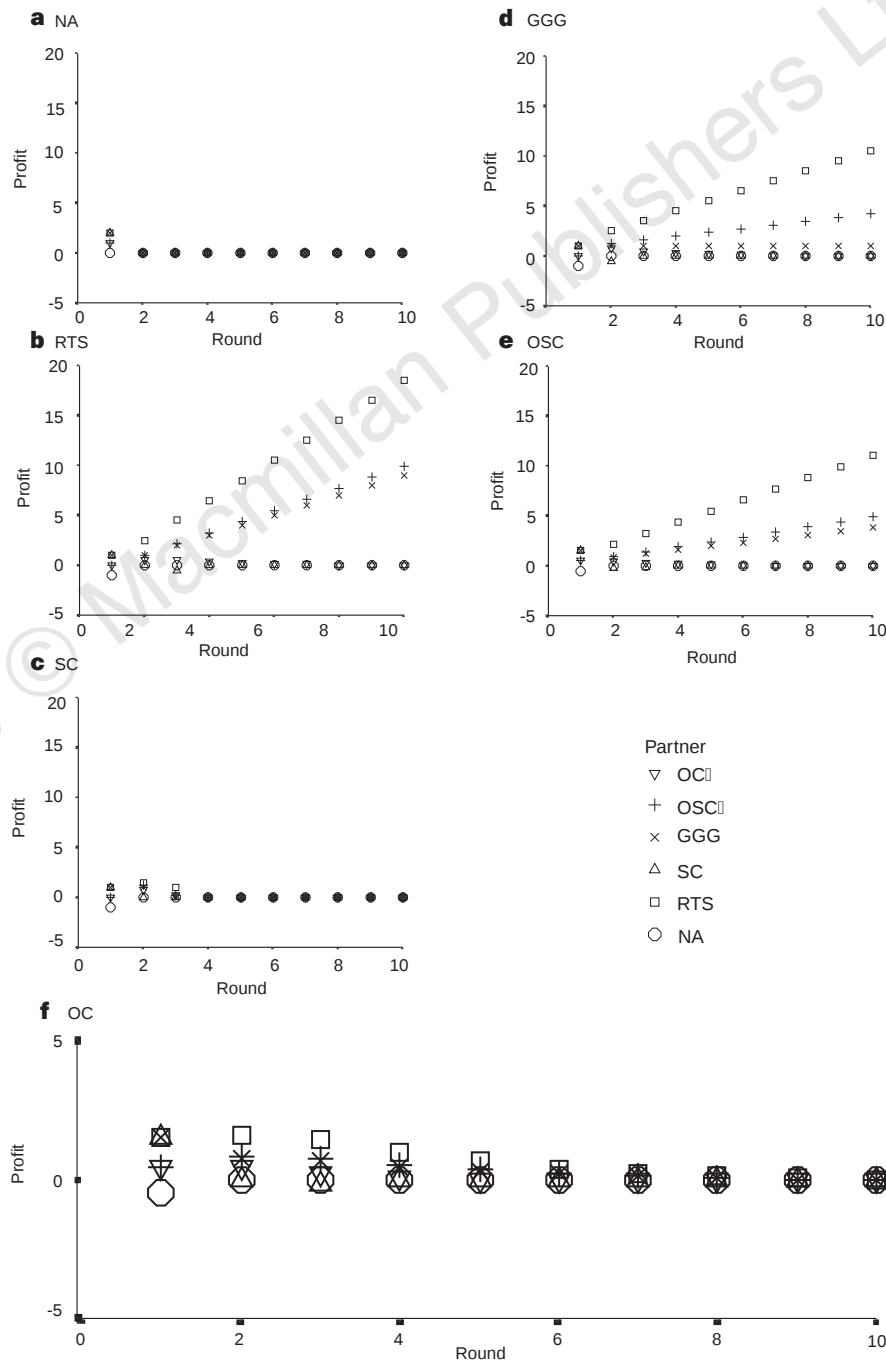


Figure 2 Profits obtained by each strategy in turn when playing against each other strategy in each round. The data shown are for the first ten rounds in the first generation from the simulations shown in Fig. 1a. Profits are expressed as means from 100 simulations and were obtained from the benefit received when the strategy was the partner, minus the cost incurred when the strategy was the focal agent performing an altruistic act. **a-f**, Profits obtained by NA, RTS, SC, GGG, OSC and OC, respectively.

pus) (ranking the number of grooms per bout within each of three longer interactions in Table 1 of ref. 7 and then pooling data across sequences, gives a correlation between partners, $r_s = 0.598$, $P = 0.018$, $n = 15$). There is also some evidence for our prediction that investment will initially be relatively low and will then increase

(for the first four bouts for each individual in the above data set, there is a correlation between the number of grooms per bout and the order of occurrence, $r = 0.523$, $P = 0.009$, $n = 24$). However, other analyses have failed to find consistent patterns (B. L. Hart, personal communication).

The prediction that reciprocal relationships will develop from small beginnings is also consistent with the observation¹¹ that, in the 'live-and-let-live' system of trench warfare, "restraint undertaken in certain hours could be extended to longer hours". More generally, the escalation of altruism shown by RTS mirrors the tendency of people to form friendships¹ and to act preferentially towards friends¹². It could also be argued that the potlatch¹³ represents an extreme case of altruistic escalation. A further prediction of our model is that behaviours that can be broken down into small units (such as grooming) will be more stable than those that cannot (such as potentially reciprocal arrangements for taking the lead in defending a territory¹⁴). If this is the case, then exchange of more costly acts could be facilitated by the establishment of reciprocity in some lesser currency. Thus, grooming could be a prelude to food-sharing¹⁵ and other alliances¹⁶.

The additional dimension of altruistic investment makes the potential range of strategies even greater than in the discrete PD model, and this is where a genetic algorithm approach (T.N.S. and G.R., manuscript in preparation) could be particularly valuable. Nevertheless, as the robustness of Axelrod and Hamilton's work² illustrates, much can be learned from simple models with intuitive strategies. Where reciprocal altruism brings net benefits to the participants, greater investment in altruism will lead to greater rewards^{17,18}. However, individuals performing a highly altruistic act without earlier experience of their partner are making a 'leap of faith' and laying themselves open to exploitation. Our RTS strategy shows how altruists can protect themselves by 'testing the water', investing a little on first meeting and then incrementally increasing investment only if it is reciprocated. Even without partner choice¹⁹, RTS players can achieve much greater benefits than non-altruists, short-changers, or occasional cheats, which cannot effectively exploit RTS players. Altruism in the form of RTS should predominate over 'averaging' strategies whenever there are cheats, subtle cheats or indeed any individuals which are, at the time, unwilling or unable to reciprocate adequately. A satisfying aspect of our model is that it represents an important step towards more biologically realistic treatments of cooperation. It should help to bridge the current gulf between theoreticians and those biologists who have questioned the degree to which reciprocity theory contributes to our understanding of cooperative behaviour^{20,21}. □

Methods

We carried out computer simulations by assigning a strategy to each of N agents. The program cycled through each pairwise combination of agents in turn, checking that they had never played each other before, and assigning one at random as the leader. The leader and follower played as the 'focal agent' in alternating sequence^{22,23}, such that they each made one move in each of R rounds. In each move, the focal agent performed an altruistic act as dictated by its strategy. The partner did not respond until it was its turn to be focal. This procedure was repeated for at least 30 generations. At the end of each generation, the number of each strategy entering the following generation was calculated in proportion to the mean relative payoffs of the strategies, and strategies gaining a mean payoff of less than zero were eliminated. The model was repeated 100 times for every set of conditions.

Received 4 February; accepted 27 April 1998.

1. Trivers, R. L. The evolution of reciprocal altruism. *Q. Rev. Biol.* **46**, 35–57 (1971).
2. Axelrod, R. & Hamilton, W. D. The evolution of cooperation. *Science* **211**, 1390–1396 (1981).
3. Nowak, M. & Sigmund, K. Tit-for-tat in heterogeneous populations. *Nature* **355**, 250–253 (1992).
4. Nowak, M. A. & Sigmund, K. A strategy of win–stay, lose–shift that outperforms tit-for-tat in the Prisoner's Dilemma game. *Nature* **364**, 56–58 (1993).
5. Wilkinson, G. S. Reciprocal food sharing in vampire bats. *Nature* **308**, 181–184 (1984).
6. Milinski, M. Tit for Tat and the evolution of cooperation in sticklebacks. *Nature* **325**, 433–435 (1987).
7. Hart, B. L. & Hart, L. Reciprocal allogrooming in impala, *Aepyceros melampus*. *Anim. Behav.* **44**, 1073–1083 (1992).

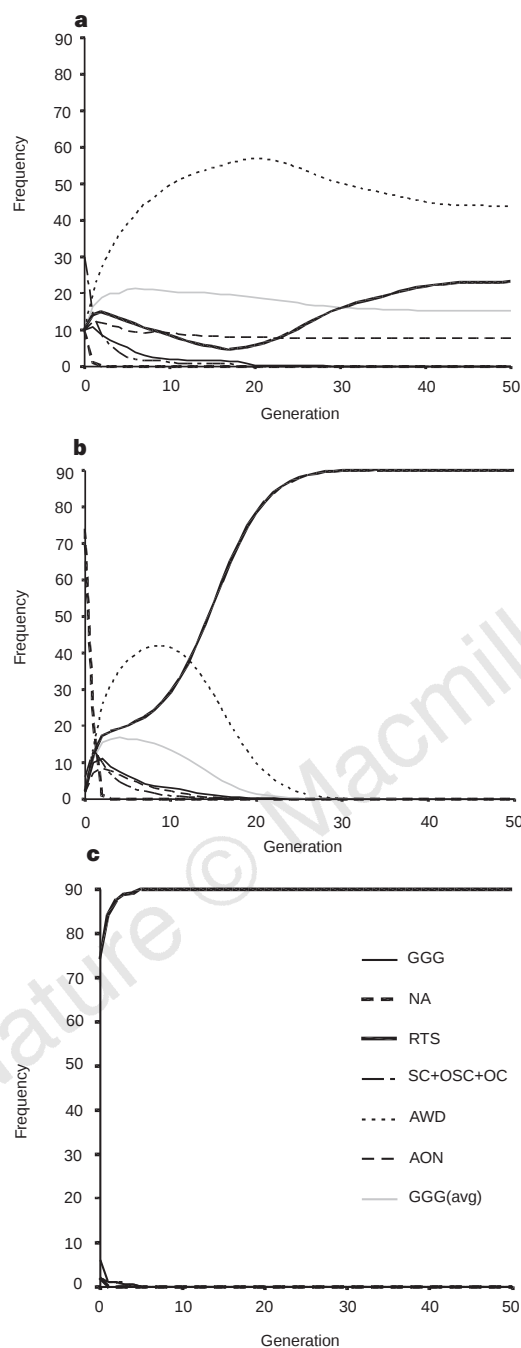


Figure 3 Here we illustrate how a strategy of increasing investment fares against strategies that invest at some average level, in competition with less altruistic strategies. In each set of simulations, $N = 90$, $R = 20$, $k = 2$ and strategy frequencies in each generation are plotted as means over 100 simulations. RTS is represented by RTS₁₃. For GGG, AWD and AON, a was defined such that over the 20 rounds, the total payoff of any of these strategies playing themselves was equal to the payoff of RTS₁₃ playing RTS₁₃. Strategies OSC and OC played as RTS₁₃ but with a probability of cheating of 0.5. **a**, At the start of these simulations, all strategies were in equal numbers. **b**, RTS invades: the initial population comprised two of each altruistic strategy, the remainder being NA. **c**, RTS resists invasion: the initial population was largely RTS and included two of each other strategy.

8. Frean, M. R. The evolution of degrees of cooperation. *J. Theor. Biol.* **182**, 549–559 (1996).
9. Mesterton-Gibbons, M. On the iterated prisoner's dilemma in a finite population. *Bull. Math. Biol.* **54**, 423–443 (1992).
10. Maynard Smith, J. *Evolution and the Theory of Games* (Cambridge Univ. Press, Cambridge, 1982).
11. Axelrod, R. M. *The Evolution of Cooperation* (Basic Books, New York, 1984).
12. Halpern, J. J. Elements of a script for friendship in transactions. *J. Conflict Res.* **41**, 835–868 (1997).
13. Benedict, R. *Patterns of Culture* (Routledge and Kegan Paul, London, 1935).
14. Heinsohn, R. & Packer, C. Complex cooperative strategies in group-territorial African lions. *Science* **269**, 1260–1263 (1995).
15. de Waal, F. B. M. The chimpanzee's service economy: food for grooming. *Evol. Hum. Behav.* **18**, 375–386 (1997).
16. Seyfarth, R. M. & Cheney, D. L. Grooming alliances and reciprocal altruism in vervet monkeys. *Nature* **308**, 541–543 (1984).
17. Roberts, G. Competitive altruism: from reciprocity to the handicap principle. *Proc. R. Soc. Lond. B* **265**, 427–431 (1998).
18. Sherratt, T. N. & Roberts, G. The evolution of generosity and choosiness in co-operative exchanges. *J. Theor. Biol.* (in the press).
19. Peck, J. R. Friendship and the evolution of co-operation. *J. Theor. Biol.* **162**, 195–228 (1993).
20. Connor, R. C. Impala allogrooming and the parcelling model of reciprocity. *Anim. Behav.* **49**, 528–530 (1995).
21. Clements, R. & Stephens, D. C. Testing models of non-kin cooperation: mutualism and the Prisoner's Dilemma. *Anim. Behav.* **50**, 527–535 (1995).
22. Frean, M. R. The prisoner's dilemma without synchrony. *Proc. R. Soc. Lond. B* **257**, 75–79 (1994).
23. Nowak, M. A. & Sigmund, K. The alternating Prisoner's Dilemma. *J. Theor. Biol.* **168**, 219–226 (1994).

Acknowledgements. We thank J. Lazarus, B. Charlton, R. Barton and M. Petrie for discussion and comments. G.R. is supported by a Lord Adams Fellowship.

Correspondence and requests for materials should be addressed to G.R. (e-mail: gilbert.roberts@ncl.ac.uk).

Visual synchrony affects binding and segmentation in perception

Marius Usher & Nick Donnelly

Department of Psychology, University of Kent at Canterbury, Kent CP2 7NP, UK

The visual system analyses information by decomposing complex objects into simple components (visual features) that are widely distributed across the cortex^{1,2}. When several objects are present simultaneously in the visual field, a mechanism is required to group (bind) together visual features that belong to each object and to separate (segment) them from features of other objects. An attractive scheme for binding visual features into a coherent percept consists of synchronizing the activity of their neural representations^{3–6}. If synchrony is important in binding, one would expect that binding and segmentation are facilitated by visual displays that are temporally manipulated to induce stimulus-dependent synchrony. Here we show that visual grouping is indeed facilitated when elements of one percept are presented at the same time as each other and are temporally separated (on a scale below the integration time of the visual system⁷) from elements of another percept or from background elements. Our results indicate that binding is due to a global mechanism of grouping caused by synchronous neural activation, and not to a local mechanism of motion computation.

Despite compelling neurophysiological support for the synchrony-binding hypothesis from animal studies (for reviews see refs 8, 9), the first psychophysical studies on humans that tested the effects of temporal manipulations on visual binding did not provide positive results^{10–12}. However, a new series of studies^{13–15} demonstrated that human subjects can perform texture discrimination when texture and background elements were spatially identical but presented in a different temporal phase (that is, discrimination was performed on the basis of temporal information only). Although this might be the first psychophysical confirmation of the use of temporal synchrony for visual binding in humans, an alternative explanation needs to be ruled out. Texture discrimination can be computed either on the basis of grouping (of the elements within each texture separately) or on the basis of local gradients at texture boundaries¹⁶. In this case, one could argue that gradient computations at boundaries generate a (possibly implicit) motion signal that

partially mediates the effect. To rule out such an explanation, we used two types of experiment that are different from texture discrimination in an attempt to, first, engage global grouping and segmentation processes but not local boundary computations, second, demonstrate the existence of a grouping mechanism that is independent of the computation of motion (even if only implicit), and third, show that grouping by temporal asynchrony interacts with spatial information, thus excluding a statistical decision based on two independent sources of information.

The first set of experiments tested grouping within a symmetric square lattice display (Fig. 1a), which is typically perceived as either rows or columns^{17,18}. Three types of display were randomly presented in a mixed design (Fig. 1a), and subjects were required to perform a forced choice about their perception of the lattice (rows or columns). In condition one, all the elements were flashed on and off together (synchronous trials); in condition two, at each successive time cycle alternating rows of the lattice were flashed on together (asynchronous-row trials); in condition three, alternating columns were flashed on together (asynchronous-column trials). The temporal asynchrony was 16 ms, which corresponds to a total time cycle of 32 ms. This asynchrony resulted in a perfectly steady lattice percept with no flicker or motion reported by subjects.

As expected, decisions in synchronous trials were evenly divided between rows and columns (because of the symmetry of the square lattice). However, during asynchronous displays the temporal structure affected subjects' perception. The probability of choosing rows or columns consistent with the temporal manipulation is shown in Fig. 1b, and is much larger than would be expected by chance (for one sample $t[7] = 35.3$, $P < 0.001$), although display times were so short that eye movements could not be made and the temporal asynchrony was much lower than that which can be detected in temporal discrimination tasks using similar displays¹⁹.

A second interesting result is that the probability of correct detection in asynchronous trials increased with contrast ($F(1, 7) = 34.76$, $P < 0.01$) but decreased with increasing display duration (Fig. 1b; $F(2, 14) = 6.84$, $P < 0.01$). Although the effect of display duration is opposite to that predicted by Bloch's law, it can be explained by a simple model showing that short displays (resulting in transient activation) are more distinct than longer displays (which engage a steady activation pattern), as shown in Fig. 1d. Finally, we found a strong trend towards better performance when using circles rather than crosses in the display ($F(1, 7) = 5.3$, $P = 0.055$). This might indicate that the impact of stimulus-induced synchrony is larger when internally induced synchronization does not dominate; internal synchronization is likely to be stronger when using crosses than when using non-orientated elements (circles) because of the lateral connections between cells with similar orientation preference within the visual cortex^{20,21}. A similar finding has been reported¹⁴ showing that the tendency to choose a target based on temporal grouping is diminished when its percept conflicts with another percept based on spatial grouping.

Although the subjects reported no motion in the lattice display, we attempted to rule out the possibility that these results are based on an implicit motion computation (vertical oscillatory 'motion' of rows and horizontal oscillatory 'motion' of columns). Another group of subjects performed an experiment in which the same displays were used but in which the subjects were required to make a

Table 1 Stimulus-response matrix for the 3AFC task

	Response R	Response C	Response S
Stimulus R	0.62 (0.07)	0.15 (0.04)	0.23 (0.04)
Stimulus C	0.15 (0.04)	0.68 (0.07)	0.18 (0.04)
Stimulus S	0.25 (0.04)	0.29 (0.05)	0.46 (0.07)
Average	0.34	0.37	0.29

Results are means $\pm$ s.e.m. of response probabilities for eight subjects. The three stimuli, R (rows), C (columns) and S (synchronous), are shown in rows, and the three responses in columns. The values are normalized to the fractions of responses generated for each stimulus, and therefore the numbers in each row sum to unity.

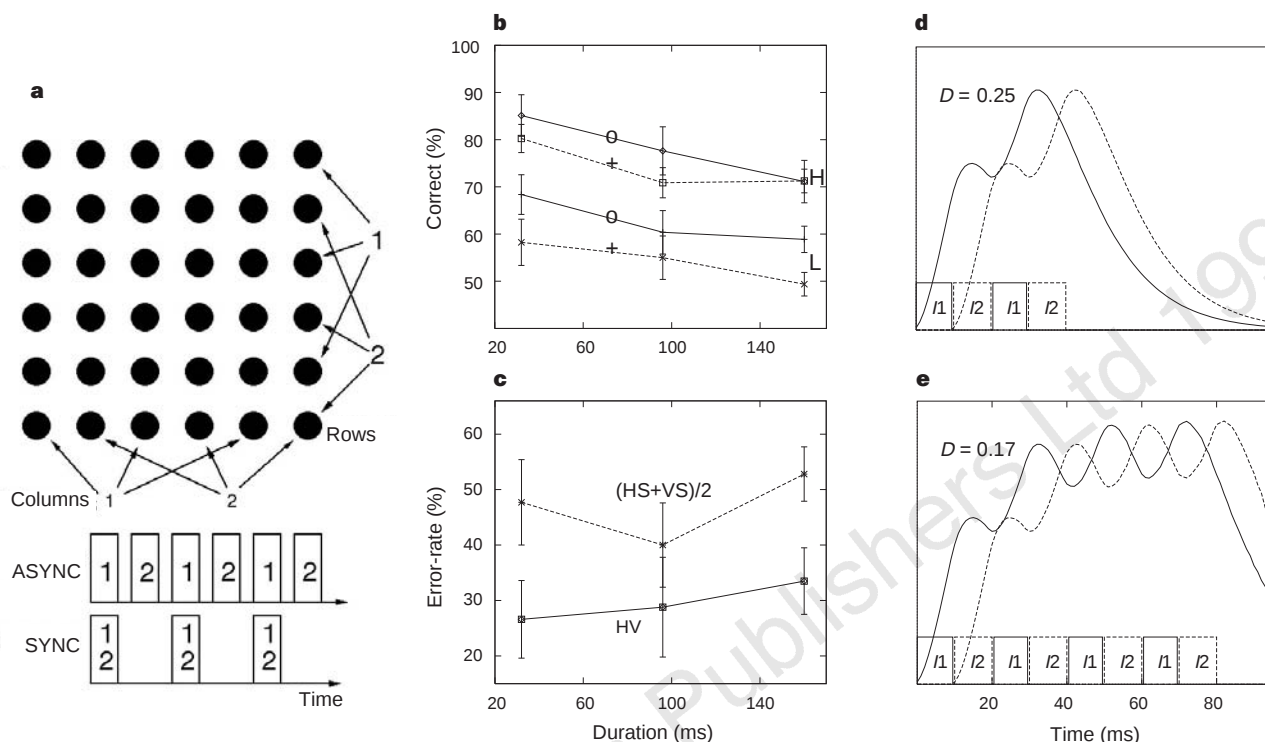


Figure 1 Effects of temporal structure on grouping. **a**, Display used in experiment 1. A square lattice of elements (either filled circles, as shown here, or crosses) was presented for a period of variable duration in one of three conditions (rows, columns and synchronous). The elements of the lattice appear relatively bright on a grey background; two levels of element brightness were used with relative contrast $((L_1 - L_2)/(L_1 + L_2))$ of 0.3 and 0.07. The temporal structure of the display involved three types of trial, randomized within a mixed design. First, synchronous trials took place in which all the elements were flashed together for a 16-ms time interval, followed by a blank screen in the following 16-ms interval. This 32-ms cycle was repeated for a period of total duration 32, 96 or 160 ms (SYNC (synchronous) condition). Second, asynchronous row trials, and third, asynchronous column trials were used. In these trials, successive rows or columns of elements, respectively, were flashed in successive 16-ms intervals (ASYNC (asynchronous) condition). Subjects were required to make a forced choice decision about their perception of the lattice. Auditory feedback was given after

three-alternative forced choice (3AFC) between row (or 'moving' vertically), column (or 'moving' horizontally) and synchronous (or 'not moving') responses. If motion was the mechanism mediating performance in this task, one would expect more confusions between types of asynchronous ('moving') display than between synchronous and asynchronous displays, as studies of motion perception have shown that the threshold for discriminating between two orthogonal directions is higher than the threshold for detecting motion²² (the thresholds are close to equal only for directions with angles larger than 120°). Subjects performed much better than chance ($P_{\text{correct}} = 0.59 \pm 0.06$ (s.e.m.); $P_{\text{chance}} = 0.33$; one sample $t(7) = 4.03$, $P < 0.005$) (Fig. 1c). A comparison of error types between conditions shows that subjects were much more likely to confuse synchronous ('non-moving') displays with either of the asynchronous ('moving') displays (HS, VS) than to confuse different asynchronous displays (HV) ($F(1, 7) = 21.28$, $P < 0.01$). Furthermore, the stimulus-response matrix (averaged across subjects and durations, Table 1) shows that the tendency to confuse synchronous and asynchronous stimuli is not the result of a moderate response bias towards generating row or column responses relative to synchronous responses overall (average response probabilities are 0.34 and 0.37 versus 0.29 respectively). When either row or column displays were presented, and errors

responses inconsistent with the stimulus manipulation. **b**, Probability of choosing rows/columns (eight subjects) consistent with the stimulus manipulation (for asynchronous trials) as a function of the total display time, two levels of contrast (L, low, or H, high) and the two types of element (circles and crosses); chance level is 50%. **c**, Error rates (eight subjects) for a triple choice between rows (horizontal, H; moving up-down, V; moving left-right, S) and static (synchronous, S). Confusions between synchronous and asynchronous-row stimuli (HS), confusions between synchronous and asynchronous-column stimuli (VS) and confusions between rows and columns (HV) are shown; chance level for errors is 66%. **d, e**, The effect of total duration on discriminability, D , of the two consecutive inputs in the ASYNC condition. Y-axis shows inputs (I_1 , I_2) and activations of hypothetical visual detectors (a_1 , a_2 , obtained by convolving the inputs with an alpha function, $f(t) = \tau \exp(-t/\tau)$, with $\tau = 10$ ms (ref. 25), for a short (**d**) and a long (**e**) display. (D is computed according to $f|a_1 - a_2|dt/f|a_1|dt$.)

made, the errors were more likely to involve synchronous responses than the opposing asynchronous response (0.23 and 0.18 relative to 0.15); this is the opposite result to that expected on the basis of a response bias alone (which should favour row/column confusions). Therefore, the pattern of confusions contradicts predictions of a motion mechanism that is better at detecting than discriminating between motions. These results indicate that grouping (in rows/columns) and not implicit motion is the mechanism responsible for the effect.

In a second type of experiment, we tested the effect of temporal asynchrony in a task that required subjects to group and detect quasicollinear line elements from a background of randomly orientated line elements in a four-alternative force choice (4AFC) task (Fig. 2a). This is not a texture discrimination task, as elements within the background are not uniformly orientated and the 'target' does not enclose a surface. Fast grouping on the basis of only spatial information has been demonstrated for such a task²³. Subjects performed much better when target and background elements were flashed cyclically and asynchronously at a time lag of 16 ms (collinear-asynchronous condition) than in the collinear-synchronous condition ($F(1, 4) = 10.6$, $P < 0.05$) (Fig. 2b). This shows that a short temporal asynchrony of 16 ms facilitates the binding of elements within the target and their segmentation from background

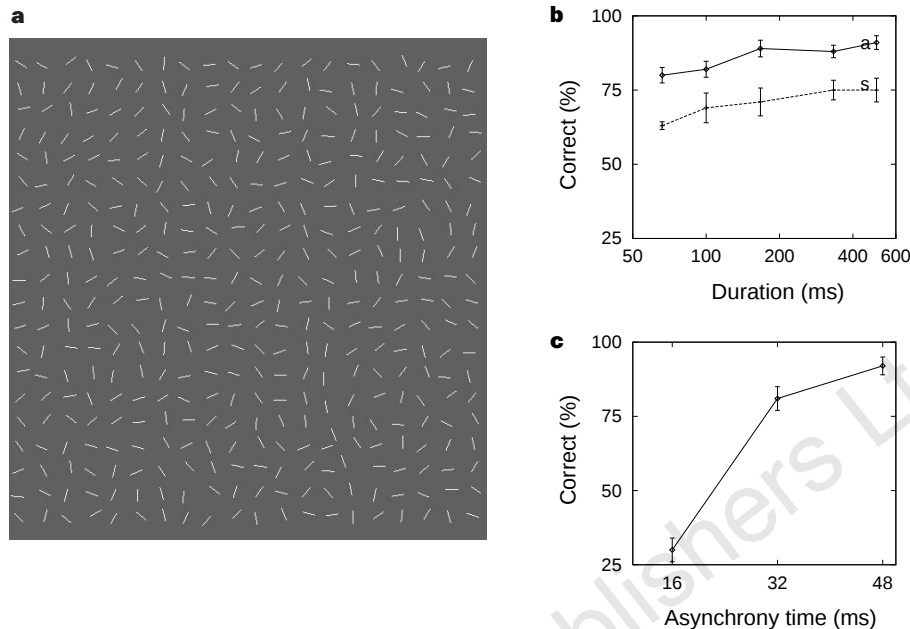


Figure 2 Effects of temporal structure on the detection of collinear target elements presented within a noisy background. **a**, Display example (the target is in the bottom right quadrant and is vertically orientated). A curved line segment is embedded within a background noise mask made of randomly orientated line elements using the algorithm described in ref. 23. The line segment is made of seven elements and appears randomly in one of the four quadrants of the display. Subjects were required to perform a 4AFC task, choosing the quadrant in which the line segment was shown. Stimuli were randomized according to the following two conditions: first, synchronous (target and background were flashed together for 16 ms, every 32 ms, as in Fig. 1a SYNC condition); and second, asynchronous (target and background were flashed in opposite phases, as in Fig. 1a ASYNC

condition). A control condition (random asynchronous) required detection for asynchronous displays when the target line elements were presented in the same positions as in the collinear condition but with their orientations randomized; under this condition the target is characterized by temporal information only. Auditory feedback was provided after incorrect responses. **b**, Probability of correct detection (five subjects) as a function of the duration of the display for two temporal structures: first, synchronous (s) and second, asynchronous (a). **c**, Probability of detecting the target against the background (ten subjects) when the orientation of the target line elements is randomized (control condition). The total display time is 200 ms, and the cycle time of the asynchrony is 16, 32 or 48 ms. In both tasks the chance level is 25%.

elements. Figure 2c shows that performance drops to chance level (25%) at the leftmost data point (one sample $t(9) = 1.83$, $P > 0.1$) when the same target elements (with the same spatial locations) have collinearity removed by randomizing their orientations so that they are distinguished from background only by the 16-ms asynchrony (random-asynchronous condition). As the random-asynchronous condition should provide the same motion signal as the original display (because the spatial-temporal difference between target and background elements is statistically the same), this shows that the facilitation effect is not due to an implicit motion detection (according to which some facilitation should have been found in the random-asynchronous condition).

A replication of this study, using a within-subject design (eight subjects) and increasing the temporal resolution of the display to 75 Hz (13 ms), produced similar results. For each subject, we compared performance in the collinear-asynchronous condition with the performance, P_{ind} , predicted from a multinomial model. This model was based on the independent probabilistic summation of spatial and temporal information and guessing (with chance probability 0.25, corresponding to 4AFC), computed according to: $P_{ind} = P_{spatial} + (1 - P_{spatial})P_{temporal} + 0.25(1 - P_{spatial})(1 - P_{temporal})$, where $P_{temporal}$ and $P_{spatial}$ are estimated from: $P_{random-async} = P_{temporal} + 0.25(1 - P_{temporal})$, $P_{collinear-sync} = P_{spatial} + 0.25(1 - P_{spatial})$. For each subject, the performance in the collinear-asynchronous condition was higher than that predicted on the basis of independent probabilistic summation; the average difference in response probability was 0.1 (one sample $t(7) = 7.32$, $P < 0.001$).

The results of the grouping in collinearity experiments indicate that the temporal information in our displays interacts with spatial information and is not used independently. This is consistent with the finding that temporal synchrony between a random set of

elements of a texture target and random elements in the background does not interfere with texture segregation on the basis of spatial features^{11,14}. A common explanation¹⁴ is that temporal grouping between elements takes place only when the target elements have compatible spatial alignments through which they can be grouped into a coherent percept (in ref. 14 grouping into a coherent percept takes place when the elements enclose a surface; in our second set of experiments it takes place when they form a continuous curve).

Our results, together with those of other studies^{13–15}, indicate that a small temporal asynchrony, below the visual integration timescale, can have a direct effect on grouping. These results are evidence for a synchrony-binding mechanism and support previous studies that showed a failure of neural synchrony to correlate with visual binding deficits in strabismic amblyopic cats²⁴. Nevertheless, further combined psychophysical and neurophysiological studies are required to test and reveal the nature of the synchrony-binding mechanism. □

Received 28 January; accepted 6 May 1998.

1. Van Essen, D. C. *Cerebral Cortex* (eds Peters, A. & Jones, E. G.) 259–329 (Plenum, New York, 1985).
2. Tanaka, K. & Saito, Y. Coding of visual images of objects in the inferotemporal cortex of macaque monkey. *J. Neurophysiol.* **66**, 170–189 (1991).
3. Milner, P. A model for visual shape recognition. *Psychol. Rev.* **81**, 521–535 (1974).
4. von der Malsburg, C. in *Models of Neural Networks II* (eds Domany, E., van Hemmen, J. L. & Shulten, K.) 95–119 (Springer, Berlin, 1981).
5. Crick, F. & Koch, C. Towards a neurobiological theory of consciousness. *Semin. Neurosci.* **2**, 263–275 (1990).
6. von der Malsburg, C. Binding in models of perception and brain function. *Curr. Opin. Neurobiol.* **5**, 520–526 (1995).
7. Coltheart, M. Iconic memory and visible persistence. *Percept. Psychophys.* **27**, 183–228 (1980).
8. Singer, W. & Gray, C. M. Visual feature integration and the temporal correlation hypothesis. *Annu. Rev. Neurosci.* **55**, 349–374 (1995).
9. Singer, W., Engel, A. K., Kreiter, A. K. & Munk, M. H. J. Neuronal assemblies: necessity, significance and detectability. *Trends Cogn. Sci.* **1**, 252–261 (1997).
10. Keele, S. W., Cohen, A., Ivry, R., Liotti, M. & Lee, P. Tests of a temporal theory of attentional binding. *J. Exp. Psych. Hum. Percept. Perform.* **14**, 444–452 (1988).

11. Kiper, D. S., Gegenfurtner, K. R. & Movshon, A. Cortical oscillatory responses do not affect visual segmentation. *Vision Res.* **36**, 539–544 (1996).
12. Fahle, M. & Koch, C. Spatial displacement, but not temporal asynchrony, destroys figural binding. *Vision Res.* **35**, 491–494 (1995).
13. Fahle, M. Figure ground discrimination for temporal information. *Proc. R. Soc. Lond. B* **254**, 199–203 (1993).
14. Leonards, U., Singer, W. & Fahle, M. The influence of temporal phase differences on texture segmentation. *Vision Res.* **36**, 2689–2697 (1996).
15. Leonards, U. & Singer, W. Two segmentation mechanisms with differential sensitivity for colour and luminance contrast. *Vision Res.* **38**, 101–109 (1998).
16. Beck, J. in *Organization and Representation in Perception* (ed. Beck, J.) 285–317 (Erlbaum, Hillsdale, NY, 1982).
17. Gregory, R. L. *Eye and Brain: The Psychology of Seeing* 3rd edn (Weidenfeld and Nicolson, London, 1997).
18. Barchilon Ben-Av, M. & Sagi, D. Perceptual grouping by similarity and proximity: experimental results can be predicted by intensity autocorrelations. *Vision Res.* **35**, 853–866 (1995).
19. Oyama, T. & Yamada, W. Perceptual grouping between successively presented stimuli and its relation to visual simultaneity and masking. *Psychol. Res.* **40**, 101–112 (1978).
20. Tso, D. Y., Gilbert, C. & Wiesel, T. N. Relationships between horizontal interactions and functional architecture in cat striate cortex as revealed by cross-correlation analysis. *J. Neurosci.* **3**, 1160–1170 (1986).
21. Lund, J. S., Takashi, Y. & Levitt, J. B. Comparison of intrinsic connectivity in different cortex of the cerebral cortex. *Cerebral Cortex* **3**, 148–162 (1993).
22. Ball, K., Sekuler, R. & Machamer, J. Detection and identification of moving targets. *Vision Res.* **23**, 229–238 (1983).
23. Field, D. J., Hayes, A. & Hess, R. F. Contour integration by the human visual system: evidence for a local “association field”. *Vision Res.* **33**, 173–193 (1993).
24. Roelfsema, P. R., Konig, P., Engel, A. K., Sireteanu, R. & Singer, W. Reduced synchronisation in the visual cortex of cats with strabismic amblyopia. *Eur. J. Neurosci.* **6**, 1645–1655 (1994).
25. Loftus, G. R., Busey, T. A. & Senders, J. W. Providing a sensory basis for models of visual information acquisition. *Percept. Psychophys.* **54**, 535–554 (1993).

Acknowledgements. We thank Y. Bonnef, E. Niebur and D. Sagi for discussions; N. Blue and A. Found for software assistance; and D. Vernon and Z. James for help in running experiments.

Correspondence and requests for materials should be addressed to M.U. (e-mail: mu@ukc.ac.uk).

Inhibitory long-term potentiation underlies auditory conditioning of goldfish escape behaviour

Yoichi Oda, Keisuke Kawasaki, Masahiro Morita, Henri Korn* & Haruko Matsui

Laboratory of Neuroscience, Division of Biophysical Engineering, Graduate School of Engineering Science, Osaka University, Machikaneyama 1-3, Toyonaka, Osaka 560-8531, Japan

* Laboratoire de Biologie Cellulaire et Moléculaire, INSERM U261, Institut Pasteur, Paris, France

Long-term potentiation (LTP), the increase in synaptic strength evoked by high-frequency stimulation, is often considered to be a cellular model for learning and memory. The validity of this model depends on the assumptions that physiological stimuli can induce LTP *in vivo* and that the resulting synaptic modifications correlate with behavioural changes. However, modifiable synapses are generally embedded deep in complex circuits. In contrast, the goldfish Mauthner (M)-cell and its afferent synapses are easily accessible for electrophysiological studies, and firing of this neuron is sufficient to trigger fast escape behaviour in response to sudden stimuli^{1,2}. We have previously shown that tetanic stimulation can induce LTP of the feedforward inhibitory synapses that control the excitability of the M-cell^{3,4}. Here we report that natural sensory stimulation can induce potentiation of this inhibitory connection that resembles the LTP induced by afferent tetanization. Furthermore, comparable acoustic stimulation produced a parallel decrease in the probability of the sound-evoked escape reflex. Thus we demonstrate for the first time, to our knowledge, a behavioural role for the long-term synaptic strengthening of inhibitory synapses.

When monosynaptic excitation of the M-cell by afferents of auditory nerve (VIII)^{1,2} causes the cell to fire, a stereotyped escape reflex, or ‘C-start’, is initiated by activation of contralateral spinal

circuits^{1,2,5}. This predator-avoidance behaviour is under the control of a feedforward glycinergic inhibition^{2,6} (Fig. 1a). We investigated whether these connections are potentiated after repeated tones and whether the C-start is similarly modified.

We first studied inhibitory postsynaptic currents (IPSCs) evoked by nerve VIII using single-electrode voltage clamp (SEVC) in the contralateral M-cell soma (Fig. 1b), where it is uncontaminated by any excitatory response. We also assessed the underlying conductance change measured from r' , which is the ratio of amplitude of the shunted (V') and control (V) antidromic spikes evoked by spinal stimulation^{4,7,8} (Fig. 1b; see Methods). Repeated sounds induced an LTP of the inhibitory synaptic current or IPSC (Fig. 1c). The IPSC evoked by nerve VIII reached $145 \pm 11\%$ ($P < 0.01$)

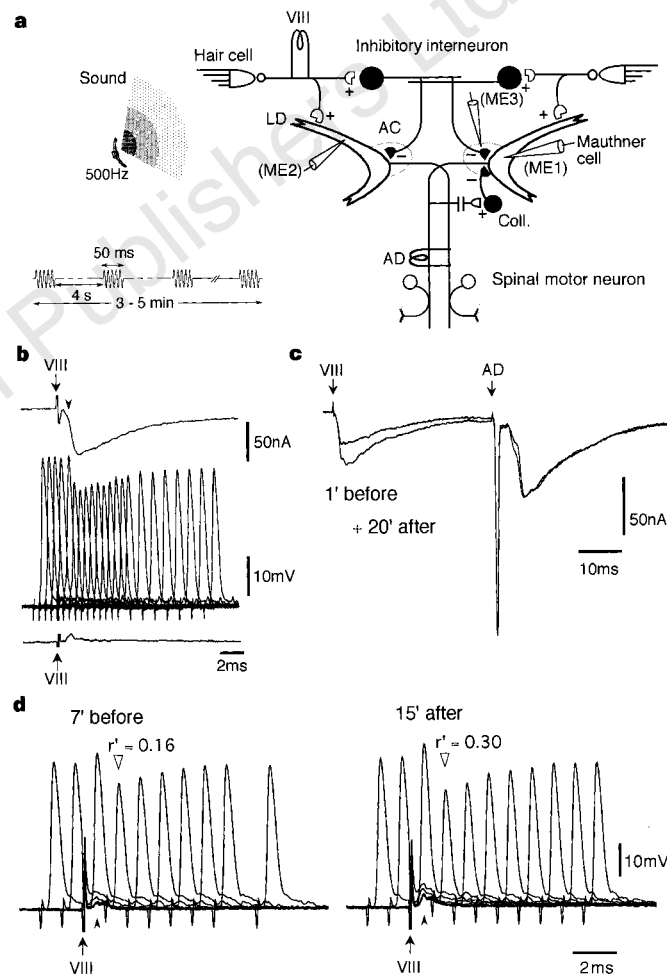


Figure 1 Induction of inhibitory LTP by sound. **a**, M-cell networks activated by sound and parameters of conditioning bursts. Nerve VIII primary fibres excite (+) the lateral dendrite (LD) of the ipsilateral M-cell and inhibitory interneurons (-). The M-cell also receives inhibition through a recurrent collateral pathway (Coll.). AD, Antidromic stimulation of the M-axon and of the recurrent pathway. ME1, ME2 and ME3: recording microelectrodes in the contralateral soma, in the ipsilateral LD, and in the axon cap (AC, dotted line). **b**, Top, average IPSC ($n = 4$) recorded at resting potential ($V_m = -81$ mV) by a single-electrode voltage clamp (SEVC), chopping frequency 20.5 KHz with a KCl microelectrode. Arrowhead, onset of IPSC. Middle, test AD spikes evoked at various intervals ($n = 3$ for each) after contralateral nerve VIII stimulation, recorded in the same cell with a KAc microelectrode. Bottom, presynaptic volley alone. **c**, IPSC evoked by nerve VIII followed by an AD spike and a subsequent collateral IPSC recorded before and after a 4 min application of repeated sounds ($n = 7$ sweeps) (SEVC, chopping frequency 18 KHz, $V_m = -77$ mV). **d**, AD spikes paired with contralateral nerve VIII stimulation as in **b**, before and after conditioning with sound bursts. Note that the time of maximum shunt (triangle) is consistent with a disinaptically evoked inhibition (diagram in **a**). Arrowhead, presynaptic volley.

of the preconditioning level, without significant change in the collateral inhibitory current ($95 \pm 5\%$, $n = 6$, $P > 0.3$).

The LTP induced by repeated tones could also be measured using r' , which is unaffected by Cl^- loading and is therefore a more stable measure of inhibitory conductance during lengthy recordings. The temporal profile of r' changed in parallel with that of the conditioned IPSC (Fig. 1d). Furthermore, the constancy of the resting membrane conductance was confirmed by the stability of the antidromic spikes (Fig. 1d).

The contralateral M-cells exhibited a sound-induced LTP of the inhibition evoked by nerve VIII (Fig. 2a), reaching $163 \pm 15\%$ of the baseline r' value ($n = 17$, $P < 0.001$), with no change in collateral inhibition ($103 \pm 2\%$, $P > 0.1$). As reported^{3,4}, the

inhibitory synapses themselves were potentiated because in 10 cells in which the presynaptic volley could be simultaneously recorded through the postsynaptic electrode, the shunt produced by a constant volley was increased to $138 \pm 4\%$ of its control value.

Unlike the contralateral M-cell, the ipsilateral M-cell receives monosynaptic mixed (chemical and electrical) excitation from nerve VIII primary afferents. In another experimental series with 16 cells, inhibitory LTP was also observed in ipsilateral M-cells (Fig. 2a, b), with r' increasing to $144 \pm 11\%$ of control values ($P < 0.002$), whereas the coupling potential (Fig. 1b) did not change significantly ($105 \pm 3\%$, $P > 0.1$) (Fig. 2a). Overall sound-induced inhibitory LTP of both M-cells lasted over 5 h.

Auditory perception in fish is optimum at frequencies in the

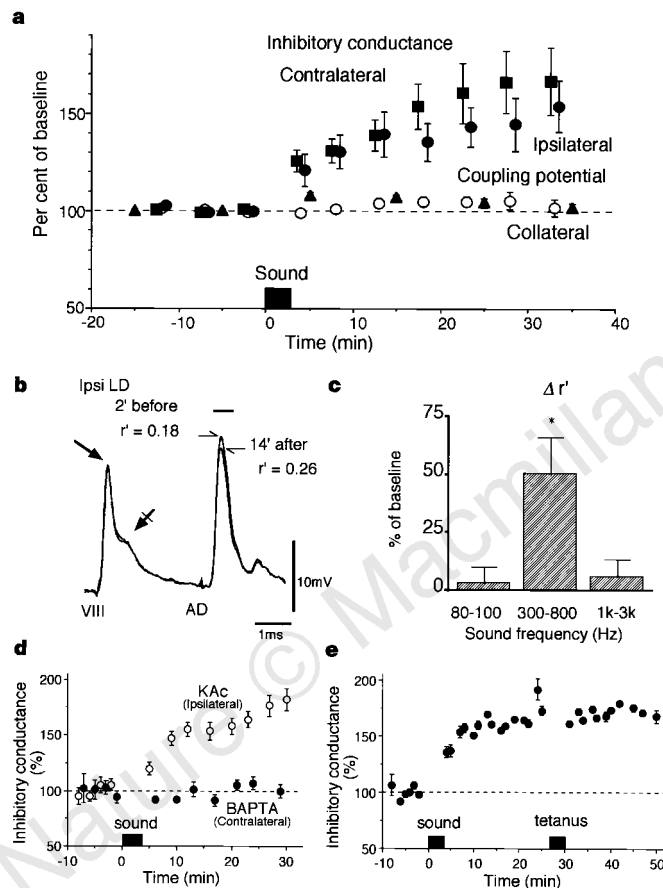


Figure 2 Properties of sound-evoked LTP. **a**, Amplitude of the inhibition evoked by nerve VIII (r') in the contralateral ($\blacksquare$, $n = 17$) and in the ipsilateral ($\bullet$, $n = 16$) M-cells of the coupling potential in the ipsilateral M-cell ($\circ$, $n = 16$) and of the contralateral collateral inhibition ($\blacktriangle$, $n = 17$), plotted versus time after onset of conditioning sound at 300–800 Hz (black rectangle). Values were normalized with respect to the preconditioning level. In some experiments ($n = 4$), 500-Hz tone conditioning was applied after control stimulation by other frequencies failed to induce LTP. **b**, Mixed EPSPs evoked by nerve VIII, with coupling potential (arrow) and chemical (crossed arrow) components, recorded in the ipsilateral LD, followed by AD test spikes evoked during the inhibitory shunt; the peak of the control spike is indicated by a horizontal bar (average of 10 sweeps). **c**, Frequency dependence of LTP. Histograms of the increase in inhibition ($\Delta r'$) onto contralateral M-cell, measured 20 min after the onset of conditioning with sounds of the indicated frequencies: 80–100 Hz ($n = 11$), 300–800 Hz ($n = 13$) and 1–3 kHz ($n = 11$). $*P < 0.008$. **d**, Plot of r' produced simultaneously in both M-cells by left nerve VIII stimulations versus time in an experiment where the contralateral M-cell was monitored with a 5 mM solution of BAPTA in the microelectrode (15 responses per data point). **e**, Plot of r' measured in an M-cell after a 3-min period of sound bursts later followed by 3 min of tetanization of the contralateral nerve VIII (10 responses per data point).

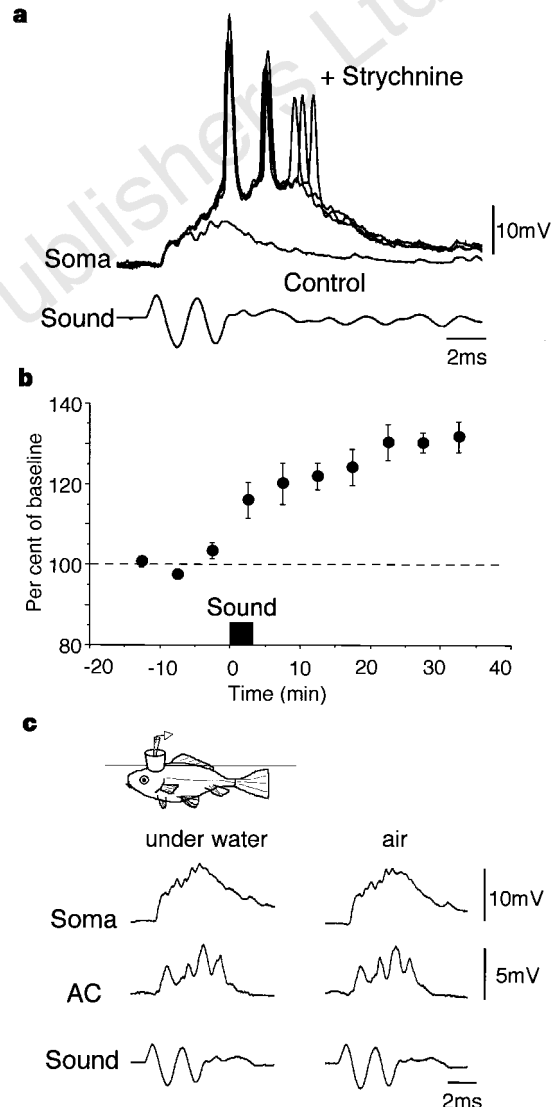


Figure 3 Activation of inhibitory pathways by sound. **a**, Depolarizing potentials evoked in the M-cell soma by a brief sound burst (lower trace) before and after surface application of 0.1 mg ml^{-1} strychnine. **b**, Plots of r' assessed 6 ms after onset of a tone burst before and after a sound-induced potentiation of inhibition versus time ($n = 11$). **c**, Sound-evoked depolarization (averages of 10 sweeps each) recorded in the M-cell soma and inhibitory presynaptic volley (or EHP) monitored in the axon cap (AC) from the same fish, placed in air (right) or in water (left), in which the animal was completely immersed with a head chamber allowing the insertion of a recording microelectrode (diagram above). In **b** and **c**, tones were of two cycles at 500 Hz and 95 dB.

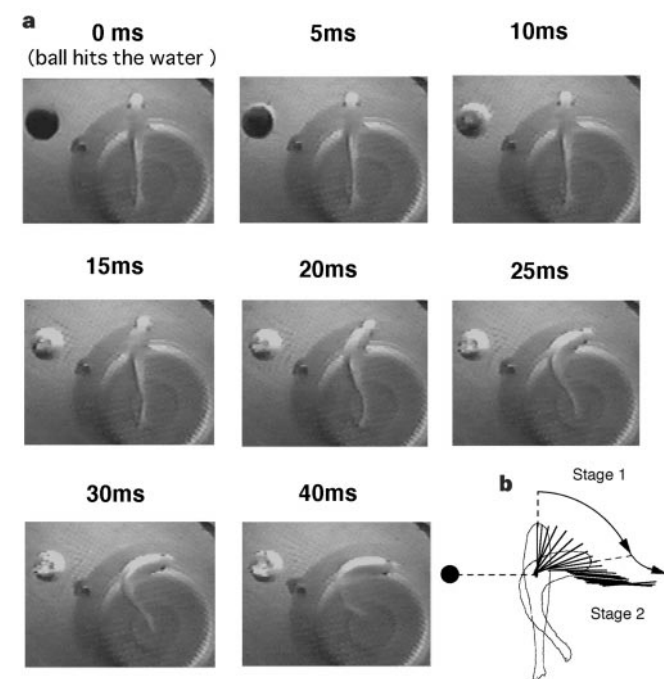


Figure 4 Components of the control motor response. **a**, Escape evoked by a falling ball, recorded at 400 frames s^{-1} . Time zero is when the ball hits the water surface. **b**, Each line denotes the position of the rostral midline, of the fish shown in **a**, from nose to centre of mass, at 2.5 ms intervals. The C-start is divided into stage 1, during which the fish bends its trunk on one side of the body, and stage 2, starting when the centre of mass begins to move (see text and ref. 11). Stages 1 and 2 are separated by a dotted line. The filled circle shows the position of the dropped ball.

range of 200 to 800 Hz^{9,10} and we found that the inhibitory LTP was maximal in this range: in the contralateral series r' increased by $51 \pm 16\%$ ($n = 13$, $P < 0.008$) for auditory bursts at 300 to 800 Hz, whereas in two other independent control groups the increase was $6 \pm 8\%$, ($n = 11$, $P > 0.4$) for frequencies of 1,000–3,000 Hz and $4 \pm 7\%$ ($n = 11$, $P > 0.5$) for 80–100 Hz (Fig. 2c).

Two types of experiment suggested that LTP induced by sound and by nerve VIII electrical stimulation^{3,4} involve similar mechanisms and pathways. First, LTP ($103 \pm 3\%$) could not be induced when a calcium chelator (BAPTA) was injected into the postsynaptic cell (Fig. 2d), whereas r' increased to $140 \pm 17\%$ in the non-injected side ($n = 3$). Second, we attempted to increase a sound-induced LTP by subsequent electrical tetanization of nerve VIII fibres at intensities known to be effective (Fig. 2e). In four of five fish, the sound induced a near-maximal potentiation ($161 \pm 19\%$, $P < 0.05$), which occluded any further increase of r' ($104 \pm 5\%$, $P > 0.4$). In the other case, a moderate sound-induced LTP of 132% was followed by an enhancement of r' to 182% by tetanization.

We next substituted acoustic for electrical test stimuli. Brief tones produced depolarizations in the M-cell, and superfusion of the brain with strychnine demonstrated (Fig. 3a) that these responses had both excitatory and inhibitory components ($n = 5$). Again the conditioning protocol enhanced the acoustically evoked inhibition with no change in the EPSPs (not shown). The conditioned r' (Fig. 3b) was enhanced by $30 \pm 3\%$ ($n = 11$, $P < 0.001$), demonstrating that inhibitory LTP does occur for synaptic responses evoked by natural sensory inputs.

As a prelude to behavioural studies, we tested whether acoustic stimuli induce similar responses in air and in water. Sound-evoked depolarizations in the M-cell and the inhibitory presynaptic volley, called extrinsic hyperpolarizing potentials (EHP) in the axon cap¹¹,

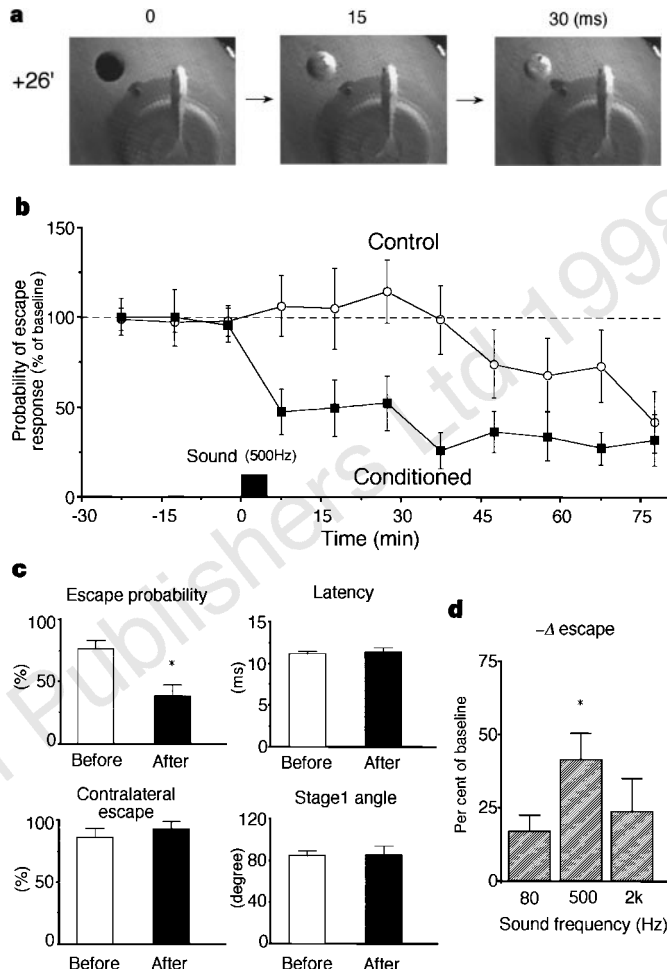


Figure 5 Behavioural effect of auditory conditioning. **a**, Sample frames obtained 26 min after repeated sound bursts (500 Hz, 94 dB; reference sound pressure = 10^{-3} Pa; ref. 28): escape was not evoked by a falling ball. **b**, Time course of behavioural changes in control and in conditioned animals (summarized from nine fish, recorded at 400 frames s^{-1} ($n = 5$) or 30 frames s^{-1} ($n = 4$)). **c**, Top left, probability of evoking an escape with a ball dropped within 60 to 160 degrees from the initial body axis (pooled data from five fish) averaged before (15 trials, 20 min) and after (20 trials, first 40 min) conditioning. * $P < 0.0001$. Top right, latency from time zero to the time at which the midline moved 4 degrees or more. Bottom left, escapes away from the ball as a percentage of total escape responses, assessed for fish initially located more than half a body length from the tank wall. Bottom right, amplitude of flexion angle during stage 1, during escapes to the contralateral side. **d**, Decrease in escape frequency ($-\Delta$ escape) tested 20 min after onset of sound stimuli delivered in water and having frequencies of 80 Hz ($n = 7$), 500 Hz ($n = 12$) and 2 kHz ($n = 7$). * $P < 0.005$.

were remarkably similar in both conditions (Fig. 3c), indicating that inhibitory LTP can occur in water.

To determine whether inhibitory LTP affects behaviour, we analysed the C-start in free-swimming fish. The C-start is initiated by short-latency (~ 2 ms) activation of one M-cell^{1,2} and consists of an initial sudden turn (stage 1) followed by a rapid acceleration (stage 2)^{12,13}. We observed these characteristics (Fig. 4a, b) in control fish. With 15 trials each of 12 fish, the frequency of escape was $67 \pm 6\%$. Conditioning with the sound subthreshold for C-start reduced this rate (Fig. 5a, b) in nine fish to less than 70% of the preconditioning level for more than 40 min, whereas it remained constant ($P > 0.1$) for at least 60 min in controls. Recovery from this depression was progressive and required at least 2 h.

Control and conditioned C-start responses were analysed

quantitatively (at 400 frames s⁻¹) in five fish (Fig. 5c) according to previous studies^{12,13}. Again the escape probability decreased after conditioning (76 ± 7% before; 39 ± 8% after; $P < 0.0001$). The less frequent C-starts had normal latencies (11 ± 0.3 ms versus 11 ± 0.5 ms; $P > 0.3$), maintained directionality away from the falling ball (85 ± 7% of positive responses versus 92 ± 6%; $P > 0.08$), and the stage-1 angle was unchanged (85 ± 5 versus 85 ± 8 degrees; $P > 0.4$). Thus, auditory conditioning did not modify the basic properties of the escape reflex.

The LTP could not be tested directly in free-swimming fish. However, the frequencies of sound effectively inducing the LTP (see Fig. 2c) were also those that had the largest effect on behaviour. Conditioning tones of 500 Hz decreased the escape probability (Fig. 5d) by 41 ± 9% of baseline ($n = 12$, $P < 0.005$), whereas, in another series, 2,000 Hz ($n = 7$, $P > 0.08$) and 80 Hz ($n = 7$, $P = 0.025$) had weaker effects (24 ± 11% and 17 ± 6%, respectively).

Recent behavioural and *in vitro* experiments have converged to suggest that LTP might be linked to modifications of complex behaviours^{14–17}. Our data strongly indicate that inhibitory LTP underlies persistent changes in a clearly delineated reflex, a conclusion supported by the similarity of: (1) M-cell electrical responses to tone bursts in air and water; and (2) the frequencies and intensities of sounds that most effectively induced cellular and behavioural modifications.

The conditioning protocol produced only subthreshold depolarizations that were sufficient to produce LTP of inhibition^{3,4} but it did not affect the M-cell's excitatory inputs¹⁸. Also nerve VIII monosynaptic EPSPs showed no long-term depression, which is produced by pairing weak afferent tetani with inhibition¹⁹. Thus inhibitory changes alone raised the firing threshold of the M-cell^{2,4}, as has been proposed for neurons that mediate crayfish escape²⁰ and *Aplysia* siphon withdrawal²¹.

Although LTP developed gradually after conditioning, the probability of escape decreased rapidly. It is possible that lateral line as well as auditory inputs converging on inhibitory interneurons²² and/or the conjunction of the test and the conditioning procedures contributed to facilitate the behavioural modifications.

The behavioural changes reported here differ from habituation of the startle reflex^{23,24} because the conditioning protocol never induced escape. Thus, this form of learning is analogous to behavioural desensitization but with a new mechanism whereby LTP of inhibition²⁵ substitutes for a depressed excitation such as that underlying habituation in *Aplysia*²⁶. This inhibitory regulation of the escape network may be critical for survival in a changeable environment (variations in depth or turbulence in noisy waters) in which a sudden motor response may become deleterious²⁷. □

Methods

Electrophysiological recording. Experiments were conducted *in vivo* on adult goldfish (*Carassius auratus*), immobilized with D-tubocurarine (1 mg per g body wt), and anaesthetized by flowing aerated water containing 70 mg l⁻¹ 3-aminobenzoic acid ethyl ester (MS 222, Sigma) through the gills. Microelectrodes containing KCl (2.8 M) or potassium acetate (KAc, 4 M) were used for voltage-clamp or current-clamp studies, as described previously^{3,4}. Test responses were evoked by stimulating the posterior branch of the left VIII nerve once per 4 or 5 s at intensities below threshold for evoking an action potential in the M-cell. In this neuron, the Cl⁻-equilibrium potential is close to the resting potential ($V_m < -75$ mV). To avoid complications introduced by non-stationarity of Cl⁻ loading, the inhibitory conductance G_{ipsp} was calculated from its shunting effect on test antidromic spikes according to $r' = (V/V' - 1)$, which equals G_{ipsp}/G_m where G_m is the resting conductance^{7,8}. This method is appropriate because spikes propagate passively into the M-cell soma and dendrites^{2,7,8,11}. We measured r' 2 ms after the nerve VIII stimulus, at the peak of the IPSC. Unless otherwise noted, conditioning stimuli were 50-ms tone bursts at 500 Hz and 90–100 dB, applied every 4 s for 3 to 5 min from a loudspeaker 25 cm from the left side of the fish. Potentiation was quantified by comparing averages 20–30 min after conditioning with those from the last

10 min of the control period. In some experiments, BAPTA (1,2-bis(o-aminophenoxy)ethane-*N,N,N',N'*-tetraacetic acid, 1.5 to 5 mM) was added to the solution of the recording electrode. Data are represented as means ± s.e.m. of all cells studied. The paired *t*-test (two-tailed) was used to assess statistical significance.

Behavioural analyses. We studied the C-start in an opaque tank, 70 cm in diameter and height, with a water depth of 15 cm. Movements were monitored with a video-camera system (400 or 30 frames s⁻¹; Photron Fastcam). Escape responses were elicited by dropping an 18 g plastic ball, 3.3 cm in diameter, from a height of 15 cm onto the water surface; the test was performed once every minute. Sessions of five test trials were each separated by 5 min. Escape frequency, turn movement, onset latency and C-start direction were measured. Turn movement is defined as the angle of body flexion during stage 1, which lasts until the centre of mass has moved 7 mm horizontally from the starting position¹². For conditioning, each animal was exposed to sound bursts (50 ms) applied every 4 s for 3–5 min by an underwater loudspeaker (University Sound, UW-30). These stimuli were subthreshold for evoking an escape response and comparable to those that induced LTP, that is, an intensity of 92–100 dB (equivalent to 40–100 Pa).

Received 19 February; accepted 20 April 1998.

- Zottoli, S. J. Correlation of the startle reflex and Mauthner cell auditory responses in unrestrained goldfish. *J. Exp. Biol.* **66**, 243–254 (1977).
- Faber, D. S. & Korn, H. *Neurobiology of the Mauthner Cell* (Raven, New York, 1978).
- Korn, H., Oda, Y. & Faber, D. S. Long-term potentiation of inhibitory circuits and synapses in the CNS. *Proc. Natl Acad. Sci. USA* **89**, 440–443 (1992).
- Oda, Y., Charpier, S., Murayama, Y., Suma, C. & Korn, H. Long-term potentiation of glycinergic inhibitory synaptic transmission. *J. Neurophysiol.* **74**, 1056–1074 (1995).
- Fetcho, J. R. Spinal network of the Mauthner cell. *Brain Behav. Evol.* **37**, 298–316 (1991).
- Zottoli, S. J. & Faber, D. S. An identifiable class of statoacoustic interneurons with bilateral projections in the goldfish medulla. *Neuroscience* **5**, 1287–1302 (1980).
- Faber, D. S. & Korn, H. Transmission at a central inhibitory synapse. I. Magnitude of unitary postsynaptic conductance change and kinetics of channel activation. *J. Neurophysiol.* **48**, 654–678 (1982).
- Faber, D. S. & Korn, H. Unitary conductance changes at teleost Mauthner cell glycinergic synapses: a voltage-clamp and pharmacological analysis. *J. Neurophysiol.* **60**, 1982–1999 (1988).
- Furukawa, T. & Ishii, Y. Neurophysiological studies on hearing in goldfish. *J. Neurophysiol.* **30**, 1377–1403 (1967).
- Fay, R. R. Physiology of primary saccular afferents of goldfish: Implications for Mauthner cell response. *Brain Behav. Evol.* **46**, 141–150 (1995).
- Furukawa, T. & Furshpan, E. J. Two inhibitory mechanisms in the Mauthner neurons of the goldfish. *J. Neurophysiol.* **26**, 140–176 (1963).
- Eaton, R. C., DiDomenico, R. & Nissano, J. Flexible body dynamics of the goldfish C-start: Implications for reticulospinal command mechanisms. *J. Neurosci.* **8**, 2758–2768 (1988).
- Eaton, R. C. & Emberley, D. S. How stimulus direction determines the trajectory of the Mauthner-initiated escape response in a teleost fish. *J. Exp. Biol.* **161**, 469–487 (1991).
- Rotenberg, A., Mayford, M., Hawkins, R. D., Kandel, E. R. & Muller, R. U. Mice expressing activated CaMKII lack low frequency LTP and do not form stable place cells in the CA1 region of the hippocampus. *Cell* **87**, 1351–1361 (1996).
- Wilson, M. A. & Tonegawa, S. Synaptic plasticity, place cells and spatial memory: study with second generation knockouts. *Trends Neurosci.* **20**, 102–106 (1997).
- Rogan, M. T., Stäubli, U. V. & LeDoux, J. E. Fear conditioning induces associative long-term potentiation in the amygdala. *Nature* **390**, 604–607 (1997).
- McKernan, M. G. & Shinnick-Gallagher, P. Fear conditioning induces a lasting potentiation of synaptic currents *in vitro*. *Nature* **390**, 607–611 (1997).
- Yang, X. D., Korn, H. & Faber, D. S. Long-term potentiation of electrotonic coupling at mixed synapses. *Nature* **348**, 542–545 (1990).
- Yang, X. D. & Faber, D. S. Initial synaptic efficacy influences induction and expression of long-term changes in transmission. *Proc. Natl Acad. Sci. USA* **88**, 4299–4303 (1991).
- Vu, E. T., Lee, S. C. & Krasne, F. B. The mechanism of tonic inhibition of crayfish escape behavior: Distal inhibition and its functional significance. *J. Neurosci.* **13**, 4379–4393 (1993).
- Fisher, T. M. & Carew, T. J. Cutaneous activation of the inhibitory L30 interneurons provides a mechanism for regulating adaptive gain control in the siphon withdrawal reflex of *Aplysia*. *J. Neurosci.* **15**, 762–773 (1995).
- Faber, D. S. & Korn, H. Inputs from the posterior lateral line nerves upon the goldfish Mauthner cell. II. Evidence that the inhibitory components are mediated by interneurons of the recurrent collateral network. *Brain Res.* **96**, 349–356 (1975).
- Rodgers, W. L., Melzack, R. & Segal, J. R. "Tail flip response" in goldfish. *J. Comp. Physiol. Psychol.* **56**, 917–923 (1963).
- Aljure, E., Day, J. W. & Bennett, M. V. L. Postsynaptic depression of Mauthner cell-mediated startle reflex, a possible contributor to habituation. *Brain Res.* **188**, 261–268 (1980).
- Groves, P. M. & Thompson, R. F. Habituation: a dual-process theory. *Psychol. Rev.* **77**, 419–450 (1970).
- Kandel, E. R. A cell-biological approach to learning. *Grass Lecture Monogr. 1* (Soc. Neurosci., Bethesda, 1978).
- Blaxter, J. H. S. in *Sensory Biology of Aquatic Animals* (eds Atema, J., Fay, R. R., Popper, A. N. & Tavolga, W. N.) 203–232 (Springer, New York, 1988).
- Canfield, J. G. & Rose, G. J. Hierarchical sensory guidance of Mauthner-mediated escape responses in goldfish (*Carassius auratus*) and Cichlids (*Haplochromis burtoni*). *Brain Behav. Evol.* **48**, 137–156 (1996).

Acknowledgements. We thank D. S. Faber, R. Miles, E. S. Ruthazer, F. Murakami, N. Yamamoto and W.-J. Song for comments. This work was supported by Grant-in-Aid for Scientific Research on Priority Areas of Japan to Y.O.

Correspondence and requests for materials should be addressed to Y.O. (e-mail: oda@bpe.es.osaka-u.ac.jp).

Cholinergic induction of network oscillations at 40 Hz in the hippocampus *in vitro*

André Fisahn, Fenella G. Pike, Eberhard H. Buhl & Ole Paulsen

MRC Anatomical Neuropharmacology Unit, University Department of Pharmacology, Mansfield Road, Oxford OX1 3TH, UK

Acetylcholine is vital for cognitive functions of the brain. Although its actions in the individual cell are known in some detail¹, its effects at the network level are poorly understood². The hippocampus, which receives a major cholinergic input from the medial septum/diagonal band³, is important in memory^{4,5} and exhibits network activity at 40 Hz during relevant behaviours⁶. Here we show that cholinergic activation is sufficient to induce 40-Hz network oscillations⁷ in the hippocampus *in vitro*. Oscillatory activity is generated spontaneously in the CA3 subfield and can persist for hours. During the oscillatory state, principal neurons fire action potentials that are phase-related to the extracellular oscillation, but each neuron fires in only a small proportion of the cycles. Both excitatory and inhibitory synaptic events participate during the network oscillation in a precise temporal pattern. These results indicate that subcortical cholinergic input can control hippocampal memory processing by inducing fast network oscillations.

To investigate the effect of cholinergic activation on network activity in the hippocampus, we used the cholinergic agonist carbachol (20 μ M), which induced fast oscillatory field activity in hippocampal slices (39 ± 3 Hz (mean $\pm$ s.d.), range 35–55 Hz; $n = 70$; Fig. 1a, b). The amplitude of the oscillation increased with increasing concentrations of carbachol up to 20–50 μ M, above which no further change was observed (0.1–200 μ M; $n = 3$; Fig. 1c). This effect was mediated by muscarinic acetylcholine receptors, as the selective muscarinic agonist muscarine (10–20 μ M; $n = 2$) mimicked the effect, and the oscillation induced by carbachol was blocked by the muscarinic antagonists atropine (10 μ M; $n = 8$), and pirenzepine (M1-subtype-selective; 10 μ M; $n = 10$; Fig. 1a). The 40-Hz oscillatory activity appeared spontaneously, was continuous over prolonged periods (up to 3 h), and was often nested in theta frequency oscillations (Fig. 1d); similar features are regularly observed during 40-Hz oscillations *in vivo*⁶, and are consistent with previous studies of cholinergically induced theta frequency oscillations *in vitro*^{8–11}.

Synchronous oscillations at 40 Hz *in vivo* are a collective behaviour of a distributed network of neurons, which cross areal boundaries¹². Cholinergically induced oscillations at 40 Hz in hippocampal slices were seen in both the CA3 and the CA1 areas (Fig. 1a). Cross-correlations between different recording sites revealed a high degree of synchrony within the CA3 area, and a phase lag of CA1 relative to CA3. These results indicate that cholinergically induced 40-Hz oscillations are generated within the CA3 area and propagate to CA1. To test this assumption, we studied hippocampal slices with cuts made to separate the hippocampal subfields. Oscillations in CA3 persisted, whereas activity was observed in neither CA1 nor dentate gyrus ($n = 4$). Therefore, the CA3 network generates intrinsically synchronized 40-Hz oscillations, and cholinergically induced 40-Hz oscillations in CA1 depend on intact connections with CA3.

Inhibitory synaptic events mediated by GABA_A (γ -aminobutyric acid A) receptors may be involved in synchronization of population activity^{7,13–15}. To test whether rhythmic inhibition is also important during cholinergically induced oscillations, we applied a GABA_A

receptor antagonist ((-)-bicuculline methochloride, 1–10 μ M). All field oscillations were blocked (Fig. 2a; $n = 7$), indicating that rhythmic inhibitory postsynaptic potentials (IPSPs) may be pivotal in phase-locking the activity of principal neurons. Furthermore, during cholinergic activation, as well as during control conditions¹⁵, minimal stimulation of GABAergic interneurons in the presence of excitatory-amino-acid blockers entrained the firing of action potentials in the pyramidal neurons (data not shown).

If IPSPs are responsible for phase-locking the oscillatory activity of pyramidal neurons, a prolongation of their duration should decrease the oscillation frequency^{7,16,17}. We tested this prediction

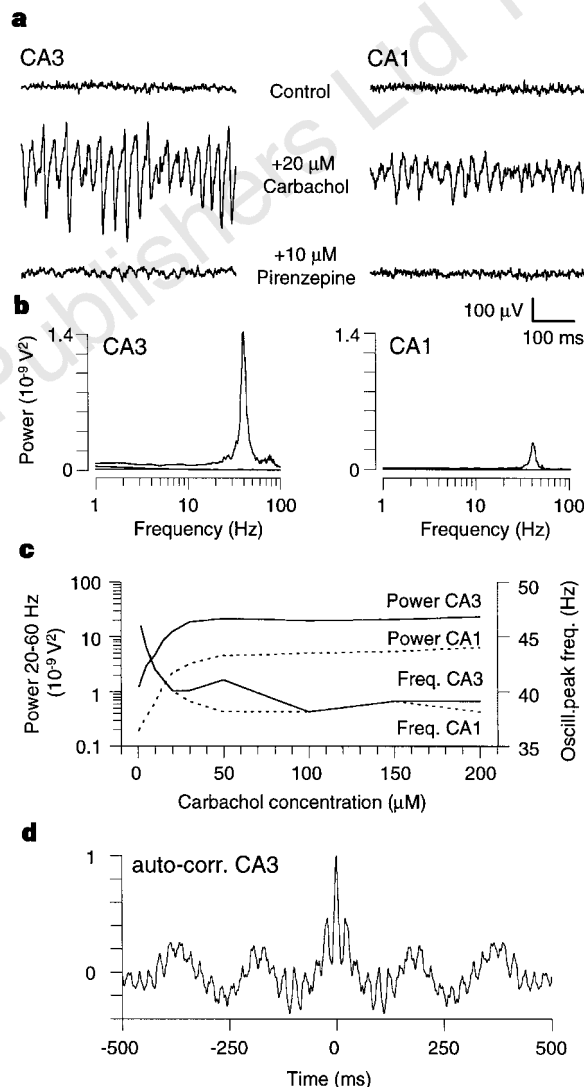


Figure 1 Generation and propagation of 40-Hz oscillations in hippocampal slices.

a, Example traces of extracellular local field recordings in the middle to distal third of the stratum radiatum of the CA3 and CA1 areas. In control solution, no apparent rhythmic activity is seen. Bath application of the cholinergic agonist carbachol induces prominent oscillatory activity which is abolished by the addition of the muscarinic M1 receptor antagonist pirenzepine. **b**, Power spectra of control, carbachol-induced extracellular oscillations and after addition of pirenzepine obtained over 50-s periods in the CA3 and CA1 areas using a fast Fourier transform algorithm. Note the distinct peak at 40 Hz in both spectra only in the presence of carbachol, whereas the spectra obtained during the control period and after pirenzepine application appear as nearly overlapping flat lines. The oscillations were consistently more prominent in the CA3 region. **c**, Concentration dependence of the carbachol-induced field oscillation. Power and frequency are plotted as a function of carbachol concentration in CA3 (continuous line) and CA1 (dotted line). **d**, Auto-correlogram of extracellular oscillations in CA3, showing periodicity at gamma and theta frequencies.

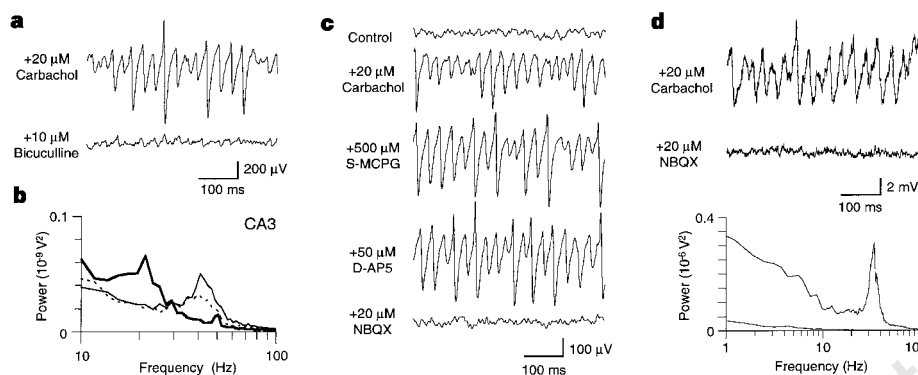


Figure 2 Pharmacological analysis of 40-Hz oscillations in the CA3 area. **a**, Extracellular field oscillations were blocked by bicuculline methochloride. **b**, Power spectra showing a reversible shift of the peak frequency of carbachol-induced oscillations following bath application of pentobarbitone. Thin line, carbachol; thick line, plus pentobarbitone; dotted line, wash. **c**, Persistence of extracellular field oscillations in the presence of the metabotropic glutamate

receptor antagonist (S)-MCPG and the NMDA receptor antagonist D-AP5. In contrast, the non-NMDA glutamate receptor antagonist NBQX blocked the oscillation. **d**, Blocking by NBQX of intracellular oscillatory activity in a pyramidal cell held at -20 mV. Power spectra in the absence (top trace) and presence (bottom trace) of NBQX are shown in the lower panel.

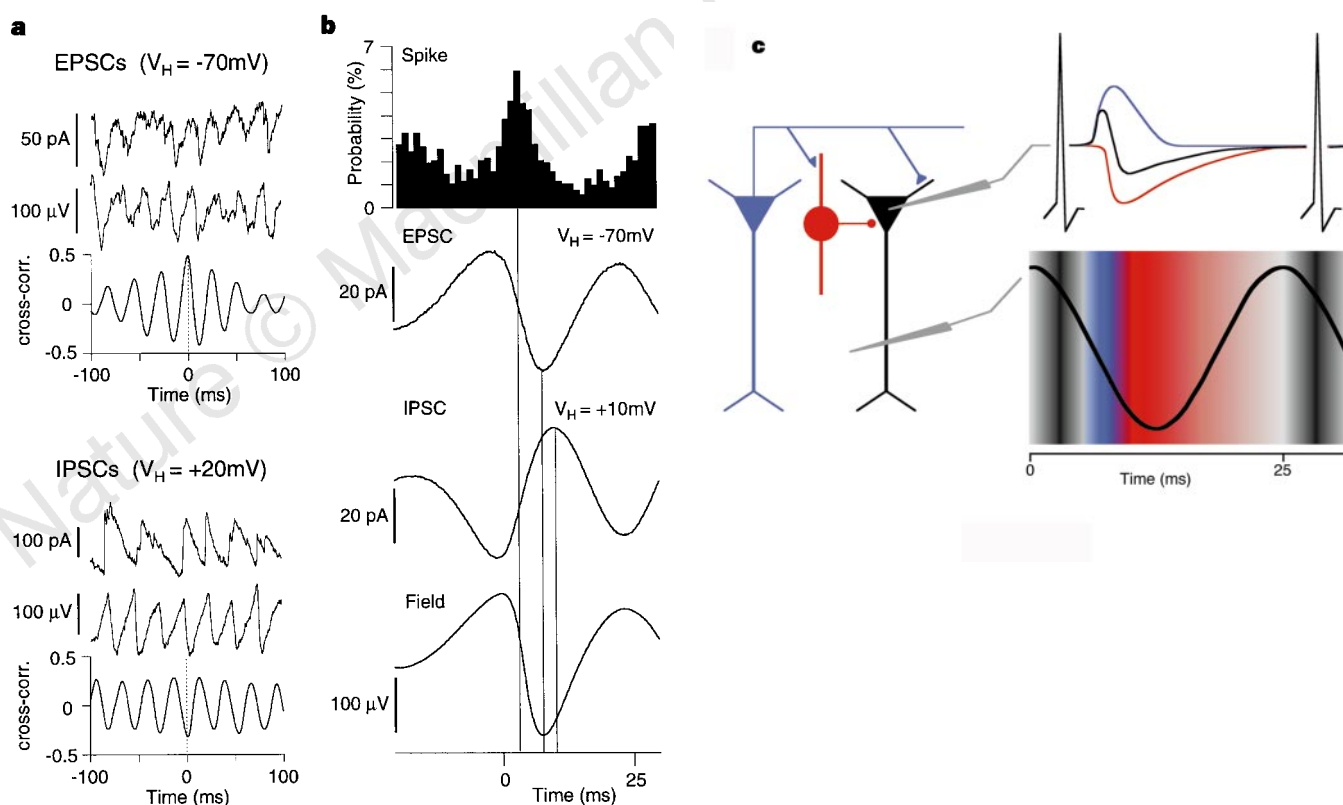


Figure 3 Temporal relationship between intracellular events and local field oscillations in CA3. **a**, Synaptic currents recorded at holding potentials of -70 mV (EPSCs) and +20 mV (IPSCs), respectively, plotted above the concomitantly recorded local field oscillation and the corresponding cross-correlogram. **b**, Occurrence of action potentials and synaptic events relative to the extracellular field oscillation. Time zero was arbitrarily set to the positive peak of the extracellular field oscillation in the stratum radiatum. Action potentials (top) are plotted as a probability histogram, showing the occurrence of all action potentials recorded over 10,000 cycles. Synaptic events (EPSC, IPSC) and field are shown as the average over 3,000 cycles. **c**, Proposed mechanism underlying the generation of 40-Hz field oscillations, based on information about the circuitry of

the CA3 region (left). Pyramidal neurons (blue) form excitatory synapses with other pyramidal neurons (black) through local axon collaterals, which also activate GABAergic interneurons (red) involved in a local inhibitory feedback circuit. Action potentials in a few pyramidal neurons therefore produce a monosynaptic EPSP (blue, top right) in neighbouring pyramidal neurons that is rapidly curtailed by a disynaptic IPSP (red, top right). The resulting summated potential is shown as a black trace (top right). Bottom right, approximate time relations between action potentials and synaptic events, relative to the field oscillation. The respective time windows, illustrating the peak probability of intracellular events, are shown in black for action potentials, blue for EPSPs and red for IPSPs.

by adding a barbiturate, which prolongs the decay time of IPSPs. Indeed, increasing the concentrations of pentobarbitone (5–20 μ M) gradually decreased the frequency of the carbachol-induced oscillation, emphasizing the role of rhythmically occurring IPSPs in pacing the activity of the hippocampal network^{7,13,15} (Fig. 2b; $n = 3$).

We next determined whether glutamatergic excitatory mechanisms are necessary for the emergent population oscillation. Metabotropic glutamate receptors are not necessary, because the receptor antagonist (S)- α -methyl-4-carboxyphenylglycine (MCPG, 200–500 μ M; $n = 6$; Fig. 2c) failed to block carbachol-induced oscillation, yet did prevent 40-Hz oscillations induced by the metabotropic glutamate receptor agonist (RS)-3,5-dihydroxyphenylglycine (DHPG, 10 μ M; $n = 2$). Likewise, NMDA (N-methyl-D-aspartate) receptors are not necessary, because the NMDA receptor antagonist D-2-amino-5-phosphonvaleric acid (D-AP5, 50 μ M; $n = 6$) failed to reduce the extracellular oscillation (Fig. 2c). However, in contrast to 40-Hz oscillations induced by metabotropic glutamate receptors in the CA1 area⁷, cholinergically induced oscillations require ionotropic non-NMDA glutamate receptors. In both CA3 and CA1 the oscillatory activity was completely abolished by the non-NMDA glutamate receptor antagonist 6-nitro-7-sulphamoylbenzo(f)-quinoxaline-2,3-dione (NBQX, 20 μ M; $n = 5$; Fig. 2c, d) as well as by the AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) receptor antagonist GYKI52466 (30–50 μ M; $n = 4$)¹⁸. Thus, tonic and/or phasic excitatory activity seems to be important for these oscillations. Recurrent excitatory feedback in the CA3 region could be involved in the generation of the oscillations, and propagation to CA1 may depend on the excitatory Schaffer collateral input onto CA1 neurons.

During 40-Hz oscillations in the hippocampus *in vivo*, individual pyramidal neurons do not fire action potentials at every cycle⁶. To analyse the frequency of action potentials and their temporal relation to synaptic events, we recorded action potentials during cell-attached patch-clamp recordings from CA3 pyramidal neurons followed by whole-cell recordings of synaptic events from the same neurons (Fig. 3a). We then temporally related both action potentials and synaptic events to extracellular field oscillations ($n = 3$; Fig. 3b). Each pyramidal neuron fired in fewer than 5% of the cycles. Most action potentials occurred 3–4 ms after the positive peak of the field oscillation and 56% of the action potentials fell within <20% of the cycle. The average maximum excitatory postsynaptic current (EPSC) occurred 3–4 ms after the action potential, whereas peak inhibitory currents reached their maximum 1–3 ms after the maximum EPSC (Fig. 3b).

The precise temporal sequence of action potentials preceding EPSCs, which, in turn, precede IPSCs, leads us to suggest the following model (Fig. 3c). Action potentials in a few pyramidal neurons produce a monosynaptic EPSP in neighbouring pyramidal neurons. This EPSP remains subthreshold in most neurons because it is rapidly curtailed by a disynaptic IPSP. This model is consistent with established effects of acetylcholine-receptor activation in the hippocampus, including excitation of pyramidal neurons^{19,20}, suppression of synaptic transmission at intrinsically originating excitatory synapses²¹ and an increase in the excitability of interneurons²². The delay between EPSPs and IPSPs is consistent with reported spike–spike latencies for unitary pyramidal/interneuron connections in the CA3 area²³. However, alternative mechanisms may also contribute: individual neurons may exhibit pacemaker-like oscillatory properties^{24,25} and cholinergic tonic excitation may foster the emergence of oscillatory activity in a network of mutually connected interneurons⁷, which, by virtue of their divergent output²⁶, phase the activity of both principal neurons and interneurons.

Although we have not studied alternative mechanisms that may be involved in the generation of 40-Hz oscillations²⁷, our data are consistent with the involvement of mutual excitatory connections between pyramidal neurons and with the existence of a fast-feed-

back inhibitory loop, suggested earlier as mechanisms involved in cortical network oscillations^{16,28}. The importance of excitatory connections²⁹ in the network oscillation reported here indicates that the synaptic strengths of specific excitatory connections in the network are instrumental in controlling which neuronal elements are active. We thus propose that cholinergic septal activation can induce an oscillatory state in the hippocampal network, in which the temporal and spatial structure of cortical (for example, entorhinal) inputs may govern information processing by selecting which subset of principal neurons leads the oscillation. Furthermore, a Hebbian type of plasticity⁵ may be important in modulating the strength of phase-related intrinsic and/or extrinsic inputs, thereby providing an effective mechanism in which information is encoded. □

Methods

We prepared horizontal 450- μ m-thick hippocampal slices from adult (sharp-microelectrode recordings) or 15–25-day-old (patch-electrode recordings) Wistar rats and maintained these slices at 34 °C at the interface between warm, humidified carbogen gas (95% O₂/5% CO₂) and artificial cerebrospinal fluid (ACSF) containing (in mM): NaCl, 126; KCl, 3; NaH₂PO₄, 1.25; MgSO₄, 2; CaCl₂, 2; NaHCO₃, 24; glucose, 10 (ref. 14). Extracellular recordings were made with glass microelectrodes containing ACSF (resistance <3 M Ω). Intracellular recordings were made with glass microelectrodes (resistance 100–120 M Ω) containing 1.5 M KCH₃SO₄, or with patch pipettes (resistance 2.5–3.5 M Ω) containing (in mM): potassium-gluconate, 100; EGTA, 0.6; MgCl₂, 5; HEPES, 40; QX-314, 5. Drugs were purchased from RBI (GYKI52466), Tocris Cookson (other amino-acid-receptor ligands) or Sigma (all other compounds). Carbachol was applied as carbamylcholine chloride. All drugs were diluted directly from frozen stock solutions into the superfusion medium. Data were recorded with Axoprobe-1A (sharp electrodes) and Axopatch-1D (patch pipettes) amplifiers and stored using a digital audio tape recorder (BioLogic DTR-1404). Data were redigitized at 10 kHz and digitally filtered at 1 kHz. For power spectrum analyses the data were filtered at 0.2 kHz with a low-pass eight-pole Bessel filter. Axograph software (Axon Instruments) was used for further analyses. For further experimental details see refs 15, 26.

Earlier studies of cholinergic effects on network activity might have been expected to uncover oscillations similar to those that we observed^{8–11}. We assume that slight differences in the preparation (horizontal versus transverse slices) or external medium (for example, potassium concentration) may account for the differences in the main frequency of oscillation observed, as it has been predicted from computer models that small changes in inhibition and excitability may significantly alter the frequency of synchronous rhythmic network activity³⁰.

Received 26 February; accepted 18 May 1998.

- McCormick, D. A. Neurotransmitter actions in the thalamus and cerebral cortex and their role in neuromodulation of thalamocortical activity. *Prog. Neurobiol.* **39**, 337–388 (1992).
- Liljenström, H. & Hasselmo, M. E. Cholinergic modulation of cortical oscillatory dynamics. *J. Neurophysiol.* **74**, 288–297 (1995).
- Shute, C. C. D. & Lewis, P. R. Cholinesterase-containing systems of the brain of the rat. *Nature* **199**, 1160–1164 (1963).
- Morris, R. G. M., Garrud, P., Rawlins, J. N. P. & O'Keefe, J. Place navigation impaired in rats with hippocampal lesions. *Nature* **297**, 681–683 (1982).
- Bliss, T. V. P. & Collingridge, G. L. A synaptic model of memory: long-term potentiation in the hippocampus. *Nature* **361**, 31–39 (1993).
- Bragin, A. et al. Gamma (40–100 Hz) oscillation in the hippocampus of the behaving rat. *J. Neurosci.* **15**, 47–60 (1995).
- Whittington, M. A., Traub, R. D. & Jefferys, J. G. R. Synchronized oscillations in interneuron networks driven by metabotropic glutamate receptor activation. *Nature* **373**, 612–615 (1995).
- Konopacki, J., MacIver, M. B., Bland, B. H. & Roth, S. H. Carbachol-induced EEG "theta" activity in hippocampal brain slices. *Brain Res.* **405**, 196–198 (1987).
- Huerta, P. T. & Lisman, J. E. Heightened synaptic plasticity of hippocampal CA1 neurons during a cholinergically induced rhythmic state. *Nature* **364**, 723–725 (1993).
- Williams, J. H. & Kauer, J. A. Properties of carbachol-induced oscillatory activity in rat hippocampus. *J. Neurophysiol.* **78**, 2631–2640 (1997).
- Fellous, J.-M., Johnston, T., Segal, M. & Lisman, J. E. in *Proc. Comput. Neurosci. Meeting, Big Sky Montana, July 1997* (ed. Bower, J. M.) (Plenum, New York, in the press).
- Engel, A. K., Kreiter, A. K., König, P. & Singer, W. Synchronization of oscillatory neuronal responses between striate and extrastriate visual cortical areas of the cat. *Proc. Natl Acad. Sci. USA* **88**, 6048–6052 (1991).
- Lytton, W. W. & Sejnowski, T. J. Simulations of cortical pyramidal neurons synchronized by inhibitory interneurons. *J. Neurophysiol.* **66**, 1059–1079 (1991).
- Buhl, E. H., Halasy, K. & Somogyi, P. Diverse sources of hippocampal unitary inhibitory postsynaptic potentials and the number of synaptic release sites. *Nature* **368**, 823–828 (1994).

15. Cobb, S. R., Buhl, E. H., Halasy, K., Paulsen, O. & Somogyi, P. Synchronization of neuronal activity in hippocampus by individual GABAergic interneurons. *Nature* **378**, 75–78 (1995).
16. Wilson, M. & Bower, J. M. Cortical oscillations and temporal interactions in a computer simulation of piriform cortex. *J. Neurophysiol.* **67**, 981–995 (1992).
17. Traub, R. D., Whittington, M. A., Colling, S. B., Buzsáki, G. & Jefferys, J. G. R. Analysis of gamma rhythms in the rat hippocampus *in vitro* and *in vivo*. *J. Physiol.* **493**, 471–484 (1996).
18. Bleakman, D. *et al.* Activity of 2,3-benzodiazepines at native rat and recombinant human glutamate receptors *in vitro*: stereospecificity and selectivity profiles. *Neuropharmacol.* **35**, 1689–1702 (1996).
19. Benardo, L. S. & Prince, D. A. Cholinergic excitation of mammalian hippocampal pyramidal cells. *Brain Res.* **249**, 315–331 (1982).
20. Cole, A. E. & Nicoll, R. A. Acetylcholine mediates a slow synaptic potential in hippocampal pyramidal cells. *Science* **221**, 1299–1301 (1983).
21. Hasselmo, M. E. & Schnell, E. Laminar selectivity of the cholinergic suppression of synaptic transmission in rat hippocampal region CA1: computational modeling and brain slice physiology. *J. Neurosci.* **14**, 3898–3914 (1994).
22. Pitler, T. A. & Alger, B. E. Cholinergic excitation of GABAergic interneurons in the rat hippocampal slice. *J. Physiol.* **450**, 127–142 (1992).
23. Miles, R. Synaptic excitation of inhibitory cells by single CA3 hippocampal pyramidal cells of the guinea-pig *in vitro*. *J. Physiol.* **428**, 61–77 (1990).
24. Llinás, R. R. The intrinsic electrophysiological properties of mammalian neurons: insights into central nervous system function. *Science* **242**, 1654–1664 (1988).
25. Gray, C. M. & McCormick, D. A. Chattering cells: superficial pyramidal neurons contributing to the generation of synchronous oscillations in the visual cortex. *Science* **274**, 109–113 (1996).
26. Cobb, S. R. *et al.* Synaptic effects of identified interneurons innervating both interneurons and pyramidal cells in the rat hippocampus. *Neuroscience* **79**, 629–648 (1997).
27. Jefferys, J. G. R., Traub, R. D. & Whittington, M. A. Neuronal networks for induced '40 Hz' rhythms. *Trends Neurosci.* **19**, 202–208 (1996).
28. Freeman, W. J. Relation between unit activity and evoked potentials in prepyriform cortex of cats. *J. Neurophysiol.* **31**, 337–348 (1968).
29. Whittington, M. A., Traub, R. D., Faulkner, H. J., Stanford, I. M. & Jefferys, J. G. R. Recurrent excitatory postsynaptic potentials induced by synchronized fast cortical oscillations. *Proc. Natl Acad. Sci. USA* **94**, 12198–12203 (1997).
30. Traub, R. D., Miles, R. & Wong, R. K. S. Model of the origin of rhythmic population oscillations in the hippocampal slice. *Science* **243**, 1319–1325 (1989).

Acknowledgements. We thank P. Somogyi and A. D. Smith for comments on an earlier version of the manuscript. E.H.B. holds a Medical Research Fellowship at Corpus Christi College, Oxford, and O.P. is the Christopher Welch Junior Research Fellow at Wadham College, Oxford. This work was supported by the British Medical Research Council, and grants from The Wellcome Trust (to O.P.) and The Royal Society (to O.P.).

Correspondence and requests for materials should be addressed to A.F. (e-mail: andre.fisahn@balliol.ox.ac.uk).

Electrical coupling underlies high-frequency oscillations in the hippocampus *in vitro*

A. Draguhn[†], R. D. Traub^{*}, D. Schmitz[†] & J. G. R. Jefferys^{*}

^{*} Department of Physiology, The Medical School, University of Birmingham, Birmingham B15 2TT, UK

[†] Institut für Physiologie der Charité, Humboldt-Universität zu Berlin, Tucholskystr. 2, 10117 Berlin, Germany

Coherent oscillations, in which ensembles of neurons fire in a repeated and synchronous manner, are thought to be important in higher brain functions. In the hippocampus, these discharges are categorized according to their frequency as theta (4–10 Hz)¹, gamma (20–80 Hz)² and high-frequency (~200 Hz)^{3–5} discharges, and they occur in relation to different behavioural states. The synaptic bases of theta and gamma rhythms have been extensively studied^{6,7} but the cellular bases for high-frequency oscillations are not understood. Here we report that high-frequency network oscillations are present in rat brain slices *in vitro*, occurring as a brief series of repetitive population spikes at 150–200 Hz in all hippocampal principal cell layers. Moreover, this synchronous activity is not mediated through the more commonly studied modes of chemical synaptic transmission, but is in fact a result of direct electrotonic coupling of neurons, most likely through gap-junctional connections. Thus high-frequency oscillations synchronize the activity of electrically coupled subsets of principal neurons within the well-documented synaptic network of the hippocampus.

The synaptic networks that generate theta and gamma rhythms depend critically on the time course of synaptic inhibition by interneurons that release γ -aminobutyric acid (GABA), which

coordinate the discharge of principal cells^{6,7}. Such mechanisms appear to be too slow⁷ to control high-frequency oscillations of ~200 Hz recorded from freely moving rats *in vivo*^{3–5}. High-frequency oscillations occur predominantly in the CA1 pyramidal layer, usually superimposed on negative 'sharp waves' lasting tens of milliseconds. High-frequency activity occurs at rest or during sleep and can be coherent over ~5 mm along hippocampal pyramidal layers^{3,4}.

We now report the occurrence of high-frequency oscillations in rat hippocampal slices *in vitro*. Low-pass filtering of extracellular field potentials at 400–500 Hz unmasked repetitive, short high-frequency discharges (Fig. 1). High-frequency oscillations occurred in 45 of 47 slices from 27 juvenile (postnatal day P18–26) rats, and in slices from two immature (P5) and four adult (>6 weeks old) rats. Autocorrelograms and power spectra revealed waveform repetition periods of 3–8 ms (125–333 Hz; Fig. 1A). Mean intra-burst frequencies were as follows: for CA1, 180 ± 47 Hz ($\pm$ s.d., $n = 33$ individual high-frequency events at independent sites in 20

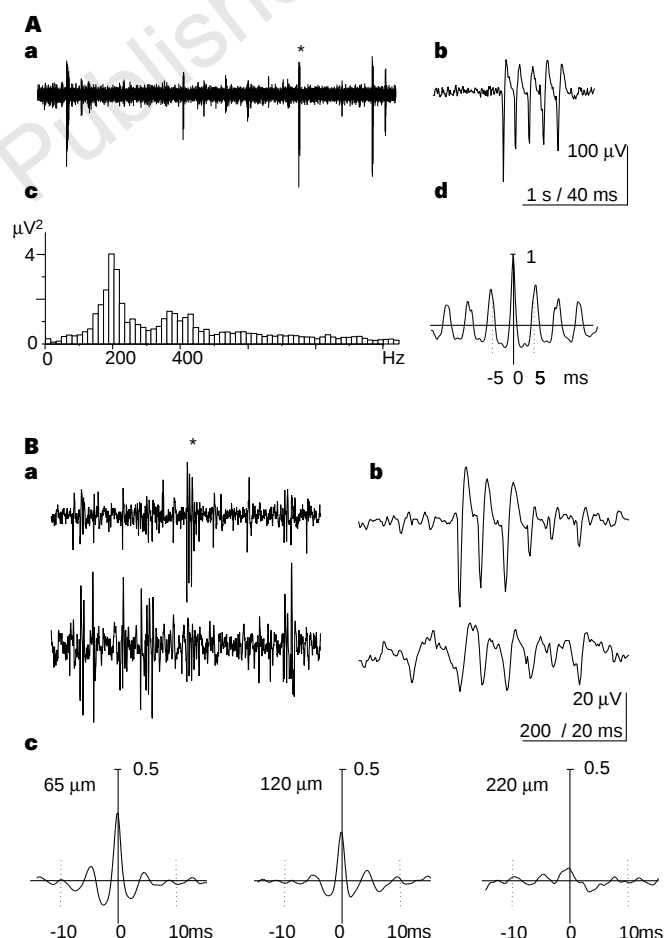


Figure 1 High-frequency oscillations in CA3. **A**, Filtered (5–400 Hz bandpass) CA3 pyramidal layer field potential with spontaneous high-frequency oscillations. **a**, **b**, The 1-s time calibration applies to the left trace (**a**); the 40-ms scale applies to the right trace (**b**, from the section marked with an asterisk in **a**). The power spectrum **c** of the trace in **a** reveals a major component at ~180 Hz. The autocorrelation function (**d**) of the marked event in **b** shows peaks at +5.4 and -5.4 ms (185 Hz). Autocorrelations from 15 high-frequency oscillations at this site yielded a mean frequency of 207 ± 18 Hz. **B**, Paired field recording from CA3 pyramidal layer at varying distances. Traces in **a** show coherent events (the section marked with an asterisk is enlarged in **b**) as well as independent events. Cross-correlograms (**c**) of 5-s periods show coherent high-frequency oscillations when the spacing between electrodes is 65 μ m or 120 μ m (phase lags 0.1 and 0.2 ms, respectively), but not when the spacing is 220 μ m.

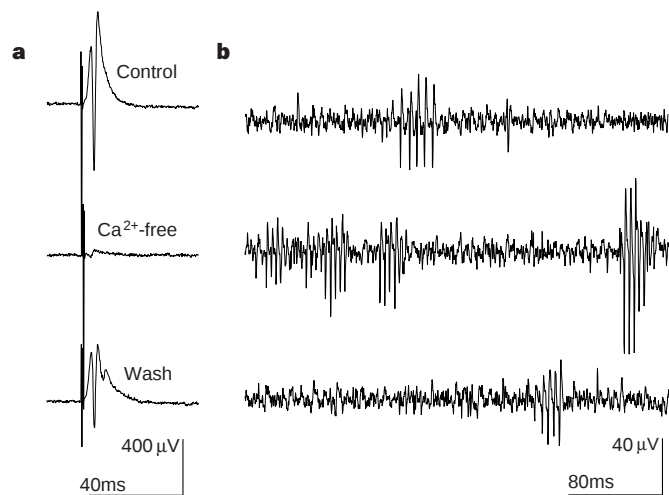


Figure 2 High-frequency oscillations do not depend on synaptic transmission. **a**, Washout of Ca^{2+} reversibly blocks field excitatory postsynaptic potentials and population spike responses recorded in the dentate granule cell layer following electrical stimulation of the perforant path. **b**, Spontaneous high-frequency activity recorded in CA3 of the same slice. When the postsynaptic response in (**a**) vanished because of washout of Ca^{2+} , high-frequency activity in CA3 increased (middle trace).

slices); for CA3, $195 \pm 44 \text{ Hz}$ ($n = 26$); and for dentate, $161 \pm 31 \text{ Hz}$ ($n = 6$). High-frequency oscillations occurred more often when neuronal excitability was enhanced by elevated K^+ concentrations (5 mM), 20 μM carbachol, 70 μM 4-aminopyridine, or 0 mM Ca^{2+} in the medium.

Laminar profiles in CA1 revealed pronounced high-frequency oscillations in the pyramidal cell layer but almost no detectable field potentials in the dendritic layers on either side ($n = 4$ of 4 slices tested). Moving the recording site along the cell layer revealed local patches of greater or less high-frequency activity.

To determine whether the high-frequency waveform reflects unit activity (that is, extracellularly recorded action potentials of single neurons) or coherent discharges in a neuronal ensemble, we inserted a second electrode at varying distances along the pyramidal layer. In four of five slices, we found clear correlations (but not identity), and consistently short phase lags, between high-frequency oscillations if the interelectrode distance was $\leq 120 \mu\text{m}$ (which corresponds to the diameters of 6–8 cells, or several hundred pyramidal cells in the slice; Fig. 1B). The presence of coherent as well as independent high-frequency oscillations at two different recording sites makes it unlikely that high-frequency field potentials reflect action potentials of a single cell.

Synchronization of high-frequency oscillations *in vivo* has been attributed to GABAergic synaptic inhibition⁴. In our experiments *in vitro*, high-frequency activity persisted after addition of the GABA_A receptor antagonist bicuculline (50 μM , $n = 13$). Further addition of ionotropic glutamate receptor antagonists (20 μM NBQX, and 40 μM D-AP5; $n = 8$) also failed to block high-frequency oscillations. To block chemical synaptic transmission completely, we used a nominally calcium-free solution. Postsynaptic field potentials ceased within 5–10 min of switching to this solution (Fig. 2a), but coherent high-frequency oscillations always became more, rather than less, prominent (9 of 9 slices, Fig. 2b). This was also the case when calcium-free artificial cerebrospinal fluid (ACSF) with 1 mM EGTA was used or when the Mg^{2+} concentration in the calcium-free ACSF was increased from 4 to 7 mM. Thus, spontaneous high-frequency oscillations persist in the absence of calcium-dependent synaptic transmission.

Synchronized 'field bursts' can occur in CA1 in the absence of synaptic transmission^{8,9} and appear to be mediated—in part—by

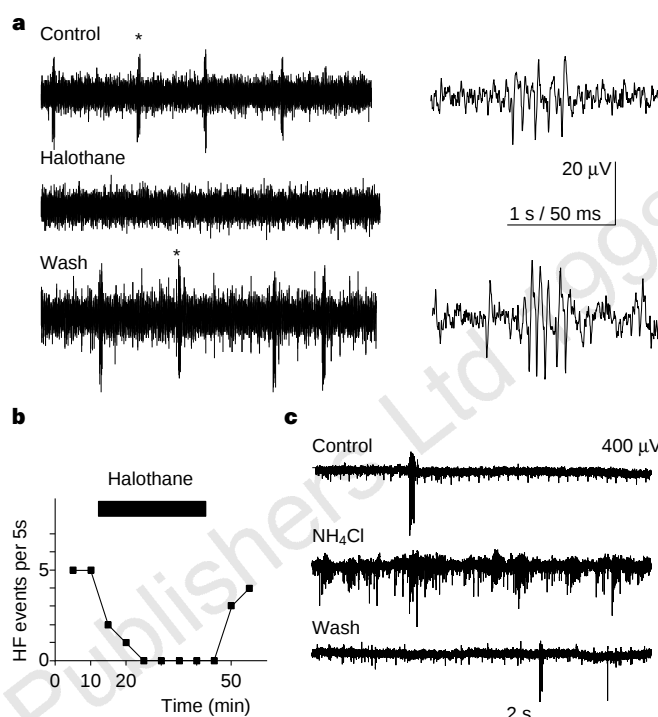


Figure 3 Gap-junctional coupling and high-frequency oscillations. **a**, High-frequency oscillations in CA3 in normal ACSF (top) and in the presence of 5 mM halothane (middle) are shown. Halothane reversibly blocks the network activity. The 1-s time calibration applies to the left traces; the 50-ms scale applies to the expanded high-frequency events at the right. **b**, Time course of the blocking of high-frequency events by 5 mM halothane. The symbols indicate the number of high-frequency events per 5-s period recorded every 5 min. **c**, Intracellular alkalization with 10 mM NH_4Cl in the ACSF increased the number and duration of high-frequency oscillations in CA3.

gap junctions^{9,10}. We therefore tested three gap-junction blockers. Octanol (1 mM, $n = 7$), halothane (5 mM, $n = 3$, Fig. 3a, b) and carbenoxolone¹¹ (100 μM , $n = 6$) all reversibly suppressed spontaneous high-frequency activity. The common action of these three agents is to block gap junctions. Halothane and carbenoxolone did not alter input resistance, time constants and discharge behaviour of CA1 neurons, but octanol reduced evoked repetitive-spike firing in 2 of 5 cells ($n \geq 5$ for each substance). Conversely, high-frequency oscillations occurred more frequently and were prolonged on adding NH_4Cl (10 mM, $n = 5$, Fig. 3c), an agent that causes intracellular alkalization¹², which enhances gap-junction coupling¹³.

To identify the cell types involved in the network oscillations, we made whole-cell current clamp recordings from 23 regular-spiking neurons in the pyramidal layer, and from 15 putative interneurons, in the vicinity of the field potential electrode. One putative CA1 pyramidal cell recorded in calcium-free solution participated unambiguously in high-frequency discharges (Fig. 4). Most negative-going population spikes within the high-frequency oscillations were accompanied by an all-or-nothing depolarization resembling 'spikelets' or 'fast prepotentials'^{10,14}. Field potential spikes remained similar whether the recorded cell exhibited subthreshold spikelets or full action potentials, indicating that the high-frequency field potential oscillations were not produced by this one cell but were generated by a multicellular network of neurons.

Computer simulations based on a detailed compartmental model of CA3 pyramidal cells¹⁵ showed that fast coupling mediated by gap junctions is possible (Fig. 4D). Realistically fast coupling potentials (or spikelets) were best reproduced when gap junctions were located

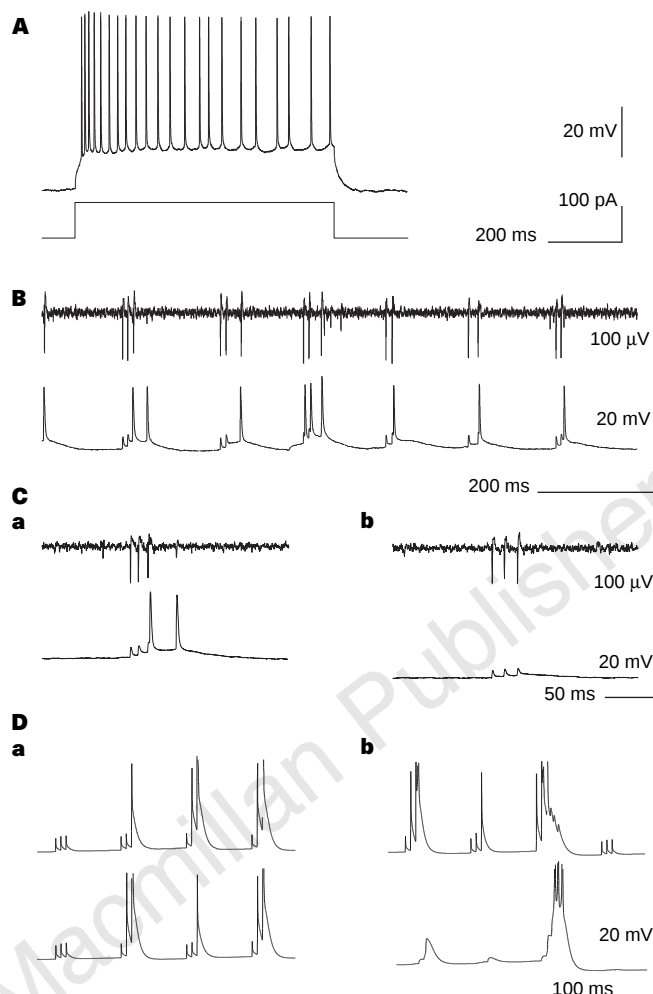


Figure 4 Cellular activity during high-frequency oscillations. **A**, Voltage response of a CA1 neuron in the principal cell layer to current injection (120 pA, whole-cell recording). The appearance of the neuron under DIC, the spike frequency adaptation, and the presence of small after-hyperpolarizations indicate that the CA1 neuron is a pyramidal cell. **B**, Parallel recording of cell potential (bottom trace) and nearby field potential (top trace; sites were about 1-cell-diameter apart) in calcium-free solution. Each field oscillation is accompanied by a depolarization and each population spike within the oscillation is accompanied by a 'spikelet' in the cell. **C**, This expanded trace shows that high-frequency activity is much larger than the field associated with isolated full action potentials. **b**, Hyperpolarization abolishes action potentials but not spikelets during high-frequency oscillations. **D**, Simulation of a pair of pyramidal cells (as described in ref. 15),

coupled either by **a**, a gap junction between axons in an unmyelinated section within 75 μ m of the initial segment, or by **b**, a gap junction between somata. The simulated gap junctions were non-rectifying and the resistance was adjusted so that some of the coupling potentials were $\sim$ 5 mV (191 M Ω in **a**; 20.5 M Ω in **b**). The distal axon of cell 1 (upper traces) was stimulated with triplets of small current pulses (period 12 ms, sufficient to induce antidromic spikes) every 150 ms. When the gap junctions are between axons (**a**), the potentials in cell 2 reflect axonal spikes that either block distal to the initial segment ($\sim$ 5 mV) or invade the soma (action potentials truncated). The smaller potentials resemble experimental records (**B**). In contrast, somatic gap junctions lead to coupling potentials in cell 2 that are smoother and slower than in experimental records.

between axons rather than between either somatic or dendritic compartments. Gap-junctional coupling between axons differs physically from coupling between somata or dendrites, as long as the density of sodium channels in the axonal membranes is high enough, because the postjunctional membrane can fire and can lead, by antidromic propagation, to the partial or full somatic action potentials observed. If the axonal sodium-channel density were low, then coupling would be largely passive and electrotonically filtered. Simulations of networks of 3,072 pyramidal cells show that an average of as few as 1.6 gap junctions per neuron can generate high-frequency oscillations without the need for chemical synapses (data not shown). The presence of small numbers of gap junctions on axons electrotonically distant from the soma may explain the lack of change in input resistance on addition of gap-junction blockers.

We conclude that spontaneous high-frequency oscillations at $\sim$ 200 Hz occur in all principal cell layers in hippocampal slices *in vitro*, are generated in small networks of pyramidal cells, and are synchronized by electrotonic coupling, probably through gap junc-

tions. Ephaptic or electric field effect interactions are unlikely to contribute greatly to synchronizing high-frequency oscillations, because the fields are smaller than the experimental estimate of the gradients (>5 mV per mm) needed to affect neuronal excitability¹⁶.

These *in vitro* high-frequency oscillations closely resemble the high-frequency 'ripples' observed in the rat hippocampus *in vivo*^{3–5}. We found a similar morphology of the events, laminar confinement to the pyramidal cell layer, participation of a small proportion of principal neurons in each 'ripple', and suppression of the oscillations by halothane. However, *in vivo*, high-frequency oscillations are coherent over $\sim$ 5 mm of CA1 pyramidal layer and are superimposed on negative waves that start in CA3 and are transmitted synaptically through CA1 to the entorhinal cortex (sharp waves)³. *In vitro*, the coherent high-frequency activity was contained within small (≤ 120 μ m) areas in physiological ACSF. 'Sharp waves' with long-range propagation of high-frequency oscillations only occurred in epileptogenic solutions, for example in the presence of bicuculline (not shown).

Gap junctions between hippocampal neurons have been shown to exist¹⁷. CA1 pyramidal cells show calcium- and pH-dependent coupling^{10,18,19} and contain connexin-43 messenger RNA²⁰, indicating that they can form gap junctions. Our results are consistent with the facts that gap-junction coupling is enhanced during intracellular alkalosis¹³ (and decreased during acidosis²¹) and that gap-junction coupling is important in synchronous bursts in calcium-free medium^{9,22}. Extracellular and intracellular potentials in CA1 pyramidal cells in calcium-free media^{10,23} resemble the phenomena observed in our experiments (Fig. 4). Our simulations of pyramidal cells, and previous simulations of interneurons²⁴, show that as few as 1.6–2.0 gap junctions per cell can lead to synchronous neuronal activity (compare with Fig. 4D). In agreement with our findings, electrotonic coupling can synchronize activity in the developing hippocampus²⁵ and generate network oscillations in sympathetic preganglionic neurons²⁶ and the inferior olive²⁷. Gap junctions can exist between axons in mammalian retina²⁸ and ectopic action potentials generated in axons need not invade the soma²⁹. Our experiments and simulations indicate that gap junctions may occur between axons in the mammalian forebrain, and may be able to synchronize neuronal activity on a fast time scale. □

Methods

We prepared transverse 400 μm hippocampal slices from Wistar rats of both sexes after anaesthesia, cervical dislocation, brain dissection and cooling to 4°C in gassed (5% CO₂, 95% O₂) ACSF containing (in mM): NaCl, 135; KCl, 3; NaH₂PO₄, 1.25; CaCl₂, 2; MgCl₂, 4; glucose, 10; and NaHCO₃, 32. Slices were maintained for 1–10 h at room temperature in a gassed, ACSF-filled storage chamber. For recording, we transferred a slice to a chamber mounted on an upright microscope with a $\times 40$ water-immersion objective and differential interference contrast (DIC) optics. The slice was superfused with ACSF at 34–35°C containing 1 mM Mg²⁺; in some experiments we used 5 mM KCl to enhance excitability. For field potential measurements we used a micro-electrode filled with HEPES-buffered solution (pH 7.4), tip broken to diameter 15–30 μm , a chlorided silver wire reference electrode less than 2 mm from the microelectrode, and differential recordings by a Neurolog 104 amplifier (Digitimer). The signal was filtered at the indicated frequencies with an active filter (Digitimer NL 125). Data were stored and analysed on a PC using SIGNAL software (CED). For whole-cell patch-clamp recordings³⁰ we used a SEC 5L (npi electronics) in the switched current clamp or bridge mode. Intracellular solution contained (in mM): KCl, 135; MgCl₂, 2; CaCl₂, 1; HEPES, 10; EGTA, 10; pH, 7.2 (KOH). NBQX (6-nitro-7-sulphamoylbenzo[f]quinoxaline-2,3-dione) and D-AP5 (D-2-amino-5-phosphonopentanoic acid) were from RBI; other drugs were from Sigma.

Received 18 March; accepted 21 May 1998.

- Vanderwolf, C. H. Hippocampal electrical activity and voluntary movement in the rat. *Electroencephalogr. Clin. Neurophysiol.* **26**, 407–418 (1969).
- Stumpf, C. The fast component in the electrical activity of rabbit's hippocampus. *Electroencephalogr. Clin. Neurophysiol.* **18**, 477–486 (1965).
- Chrobak, J. J. & Buzsáki, G. High-frequency oscillations in the output network of the hippocampal entorhinal axis of the freely behaving rat. *J. Neurosci.* **16**, 3056–3066 (1996).
- Ylinen, A. et al. Sharp wave-associated high-frequency oscillation (200 Hz) in the intact hippocampus: network and intracellular mechanisms. *J. Neurosci.* **15**, 30–46 (1995).
- Buzsáki, G., Horváth, Z., Urioste, R., Hetke, J. & Wise, K. High-frequency network oscillation in the hippocampus. *Science* **256**, 1025–1027 (1992).
- Cobb, S. R., Buhl, E. H., Halasy, K., Paulsen, O. & Somogyi, P. Synchronization of neuronal activity in hippocampus by individual GABAergic interneurons. *Nature* **378**, 612–615 (1995).
- Whittington, M. A., Traub, R. D. & Jefferys, J. G. R. Synchronized oscillations in interneuron networks driven by metabotropic glutamate receptor activation. *Nature* **373**, 612–615 (1995).
- Jefferys, J. G. R. & Haas, H. L. Synchronized bursting of CA1 pyramidal cells in the absence of synaptic transmission. *Nature* **300**, 448–450 (1982).
- Taylor, C. P. & Dudek, F. E. Synchronous neural afterdischarges in rat hippocampal slices without active chemical synapses. *Science* **218**, 810–812 (1982).
- Valiante, T. A., Perez-Velazquez, J. L., Jahromi, S. S. & Carlen, P. L. Coupling potentials in A1 neurons during calcium-free-induced field burst activity. *J. Neurosci.* **15**, 6946–6956 (1995).
- Leslie, J., Nolan, M. F., Logan, S. D. & Spanswick, D. Actions of carbenoxolone on rat sympathetic preganglionic neurones *in vitro*. *J. Physiol. (Lond.)* **506P**, 146 (1998).
- Thomas, R. C. Experimental displacement of intracellular pH and the mechanism of its subsequent recovery. *J. Physiol. (Lond.)* **354**, 3–22 (1984).
- Spray, D. C., Harris, A. L. & Bennett, M. V. L. Gap junctional conductance is a simple and sensitive function of intracellular pH. *Science* **211**, 712–715 (1981).
- Spencer, W. A. & Kandel, E. R. Electrophysiology of hippocampal neurons. IV. Fast prepotentials. *J. Neurophysiol.* **24**, 272–285 (1961).
- Traub, R. D., Jefferys, J. G. R., Miles, R., Whittington, M. A. & Tóth, K. A branching dendritic model of a rodent CA3 pyramidal neurone. *J. Physiol. (Lond.)* **481**, 79–95 (1994).

- Jefferys, J. G. R. Influence of electric fields on the excitability of granule cells in guinea-pig hippocampal slices. *J. Physiol. (Lond.)* **319**, 143–152 (1981).
- Rash, J. et al. Grid-mapped freeze-fracture analysis of gap junctions in gray and white matter of adult rat central nervous system, with evidence for a "pan-glial syncytium" that is not coupled to neurons. *J. Comp. Neurol.* **388**, 265–292 (1997).
- Church, J. & Baimbridge, K. G. Exposure to high-pH medium increases the incidence and extent of dye coupling between rat hippocampal CA1 pyramidal neurons *in vitro*. *J. Neurosci.* **11**, 3289–3295 (1991).
- MacVicar, B. A. & Dudek, F. E. Electrotonic coupling between pyramidal cells: a direct demonstration in rat hippocampal slices. *Science* **213**, 782–785 (1981).
- Simbürger, E., Stang, A., Kremer, M. & Dermietzel, R. Expression of connexin 43 in adult rodent brain. *Histochem. Cell Biol.* **107**, 127–137 (1997).
- MacVicar, B. A. & Jahnson, H. Uncoupling of CA3 pyramidal neurons by propionate. *Brain Res.* **330**, 141–145 (1985).
- Perez-Velazquez, J. L., Valiante, T. A. & Carlen, P. L. Modulation of gap junctional mechanisms during calcium-free induced field burst activity: a possible role for electrotonic coupling in epileptogenesis. *J. Neurosci.* **14**, 4308–4317 (1994).
- Patrylo, P. R., Kuhn, A. J., Schweitzer, J. S. & Dudek, F. E. Multiple-unit recordings during slow field-potential shifts in low-[Ca²⁺]₀ solutions in rat hippocampal and cortical slices. *Neuroscience* **74**, 107–118 (1996).
- Traub, R. D. Model of synchronized population bursts in electrically coupled interneurons containing active dendritic conductances. *J. Comp. Neurosci.* **2**, 283–289 (1995).
- Strata, F. et al. Pacemaker current in dye-coupled hilar interneurons contributes to the generation of giant GABAergic potentials in developing hippocampus. *J. Neurosci.* **17**, 1435–1446 (1997).
- Logan, S. D., Pickering, A. E., Gibson, I. C., Nolan, M. F. & Spanswick, D. Electrotonic coupling between rat sympathetic preganglionic neurones *in vitro*. *J. Physiol. (Lond.)* **495**, 491–502 (1996).
- Llinás, R., Baker, R. & Sotelo, C. Electrotonic coupling between neurons in cat inferior olive. *J. Neurophysiol.* **37**, 560–571 (1974).
- Vaney, D. I. The coupling pattern of axon-bearing horizontal cells in the mammalian retina. *Proc. R. Soc. Lond. B* **252**, 93–101 (1993).
- Pinaut, D. Backpropagation of action potentials generated at ectopic axonal loci: hypothesis that axon terminals integrate local environmental signals. *Brain Res. Brain Res. Rev.* **21**, 42–92 (1995).
- Hamill, O. P., Marty, A., Neher, E., Sakmann, B. & Sigworth, F. J. Improved patch-clamp techniques for high-resolution recording from cells and cell-free membrane patches. *Pflügers Arch.* **391**, 85–81 (1981).

Acknowledgements. This work was supported by a Training grant from the European Commission (to A.D.) and by the Wellcome Trust. R.D.T. is a Wellcome Principal Research Fellow. We thank U. Heinemann, H.-J. Gabriel, S. Gabriel and U. Endermann for discussions.

Correspondence and requests for materials should be addressed to A.D. (draguhn@rz.charite.hu-berlin.de).

Visualizing secretion and synaptic transmission with pH-sensitive green fluorescent proteins

Gero Miesenböck, Dino A. De Angelis & James E. Rothman

Cellular Biochemistry and Biophysics Program, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, New York, New York 10021, USA

In neural systems, information is often carried by ensembles of cells rather than by individual units. Optical indicators¹ provide a powerful means to reveal such distributed activity, particularly when protein-based and encodable in DNA^{2–4}: encodable probes can be introduced into cells, tissues, or transgenic organisms by genetic manipulation, selectively expressed in anatomically or functionally defined groups of cells, and, ideally, recorded *in situ*, without a requirement for exogenous cofactors. Here we describe sensors for secretion and neurotransmission that fulfil these criteria. We have developed pH-sensitive mutants of green fluorescent protein ('pHluorins') by structure-directed combinatorial mutagenesis, with the aim of exploiting the acidic pH inside secretory vesicles^{5,6} to monitor vesicle exocytosis and recycling. When linked to a vesicle membrane protein, pHluorins were sorted to secretory and synaptic vesicles and reported transmission at individual synaptic boutons, as well as secretion and fusion pore 'flicker' of single secretory granules.

Wild-type green fluorescent protein (GFP) has a bimodal excitation spectrum^{7,8} with peaks at 395 and 475 nm (Fig. 1a). Underlying the two maxima are protonated and deprotonated states of Tyr 66, which forms part of the chromophore^{9–13}. Although critical residues exchange protons with the environment¹⁰, the excitation spectrum of GFP is essentially unaltered between pH 5.5 and 10 (ref. 7 and

Fig. 1a). This implies that protonation–deprotonation reactions are constrained; a given GFP is trapped in either of two alternative conformations: in one the chromophore is protonated and can be excited at 395 nm, and in the other it is deprotonated and can be excited at 475 nm.

We surmised that converting GFP to a pH sensor would require amino-acid substitutions that facilitate switching between conformers, or that couple changes in bulk pH to changes in the electrostatic environment of the chromophore. To obtain such substitutions, we focused on seven key positions (Gln 94, Arg 96, His 148, Ile 167, Thr 203, Ser 205 and Glu 222) known from X-ray crystallography^{11–13} to be part of the proton-relay network of Tyr 66 or to alter the excitation spectrum when mutated^{9,11,14,15}.

Because histidine has the desired pK_a for a sensor for exocytosis, we introduced histidines at positions flanking key residues. One such mutant, S202H, showed a 16% reduction in 395-nm excitation with a 26% increase in 475-nm excitation, induced by a pH shift from 7.4 to 6.0. S202H provided the starting point for a combinatorial enhancement of pH sensitivity. Complementary DNAs of S202H (or of intermediates obtained in later rounds) were amplified by using primers in a polymerase chain reaction (PCR) that simultaneously altered between one and five codons and encoded between 20 and 8,000 different polypeptide sequences. The GFP libraries were expressed in *Escherichia coli* and a total of ~19,000 fluorescent colonies screened for pH sensitivity (see Methods). Several rounds of mutagenesis generated two classes of pH-sensitive fluorescent proteins ('pHluorins'), which we term 'ratiometric' and 'ecliptic'.

Ratiometric pHluorin contains the original mutation S202H plus E132D, S147E, N149L, N164I, K166Q, I167V, R168H and L220F.

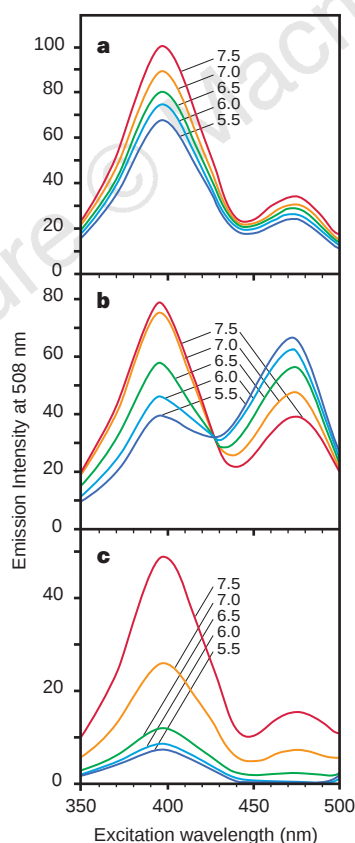


Figure 1 Fluorescence excitation spectra. **a**, Wild-type GFP; **b**, ratiometric pHluorin; and **c**, ecliptic pHluorin. Samples contained 27.5 μ M chromophore in 50 mM sodium cacodylate plus 50 mM sodium acetate, adjusted to the indicated pH values, and 100 mM NaCl, 1 mM $CaCl_2$, 1 mM $MgCl_2$. The ordinate scales reflect normalized differences in emitted fluorescence intensity.

The protein displays a reversible excitation ratio change between pH 7.5 and 5.5 (Fig. 1b), with a response time of <20 ms (not shown). Ecliptic pHluorin, by contrast, gradually loses fluorescence as pH is lowered, until at pH values of <6.0, the excitation peak at 475 nm vanishes (Fig. 1c). In an environment of pH <6.0, the protein is therefore invisible (eclipsed) under 475-nm excitation but can still be weakly seen at 395 nm. These changes are reversible within <20 ms after returning to neutral pH (not shown). Ecliptic pHluorin no longer contains S202H but carries the substitutions S147D, N149Q, T161I, S202F, Q204T and A206T. For both pHluorins, neither the minimal set of mutations nor the mechanism by which they confer pH sensitivity has been delineated. However, association of monomers into dimers^{8,12} or dissociation of existing dimers do not seem to be involved (results not shown).

For initial imaging experiments, ratiometric pHluorins were glycosylphosphatidylinositol (GPI)-anchored¹⁶ at the cell surface (Fig. 2b), inserted into the luminal domain of TGN38 (ref. 17), a membrane protein of the *trans*-Golgi network (TGN) (Fig. 2c), or attached to the lumenally exposed carboxy terminus of cellubrevin¹⁸, an endosomal membrane protein (Fig. 2d). Correct targeting of these constructs was confirmed by co-localization with standard markers (results not shown). The labelled sub-cellular compartments of transfected HeLa cells were readily seen by wide-field fluorescence microscopy and characterized by distinct 410/470-nm excitation ratios, $R_{410/470}$ (Fig. 2). The ratios remained stable within $\pm 3\%$ during 2 min of continuous image acquisition at 1 Hz, indicating lack of photoisomerization¹³.

Table 1 Ratiometric pHluorin as an indicator of synaptic transmission

Nerve terminal	$R_{410/470}$	r.m.s. noise
Resting ($n = 18$)	0.418 ± 0.021	0.016 ± 0.003
Depolarized		
+ Ca^{2+} ($n = 18$)	$0.479 \pm 0.020^*$	$0.023 \pm 0.005^*$
- Ca^{2+} ($n = 18$)	0.416 ± 0.017	0.016 ± 0.005
+BoNT B ($n = 18$)	0.421 ± 0.021	0.017 ± 0.003

$R_{410/470}$ and r.m.s. noise ($\pm$ s.d.) in recordings from individual synaptic boutons, obtained at 1 Hz. Evaluation windows extended over 10 s, starting at 3 s after application of a depolarizing stimulus. Significant differences ($P < 0.01$) with respect to resting conditions are indicated by asterisks.

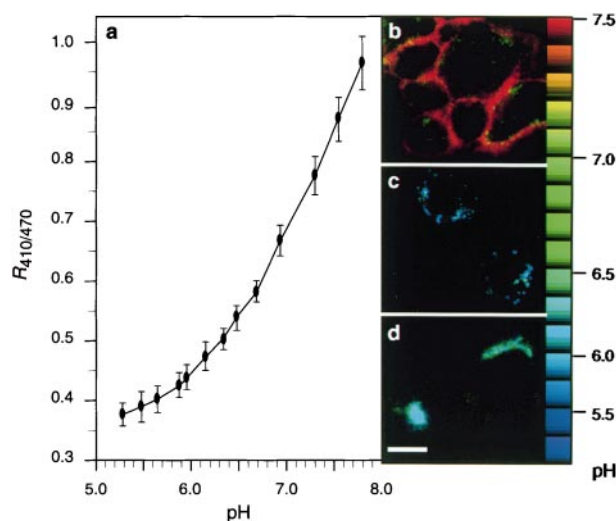


Figure 2 pH measurements with ratiometric pHluorin in HeLa cells. **a**, Calibration curve of $R_{410/470}$ (mean $\pm$ s.d.). Cells expressing GPI-anchored ratiometric pHluorin at their surface ($n = 28$) were imaged in buffers adjusted to pH values between 5.28 and 7.8. **b–d**, Ratiometric pH measurements in subcellular compartments, colour-encoded according to the look-up table on the right. Targeting modules used: **b**, GPI anchor for delivery to the cell surface (at pH 7.40); **c**, cellubrevin for sorting to endosomes; **d**, TGN38 for localization to the *trans*-Golgi network. Scale bar, 10 μ m.

To convert excitation ratios to pH values, cells expressing GPI-anchored pHluorin at their surface were imaged in buffers of defined pH. Based on this standard curve (Fig. 2a), we determined pH values (mean $\pm$ s.d.) of 5.51 ± 0.66 for endosomes ($n = 61$) and 6.21 ± 0.39 for the TGN ($n = 28$), consistent with previous estimates⁵.

Fusion of an acidified vesicle^{5,6} with the plasma membrane will establish continuity of its interior with the extracellular fluid and cause an instantaneous rise of pH to ~ 7.4 . A pHluorin,

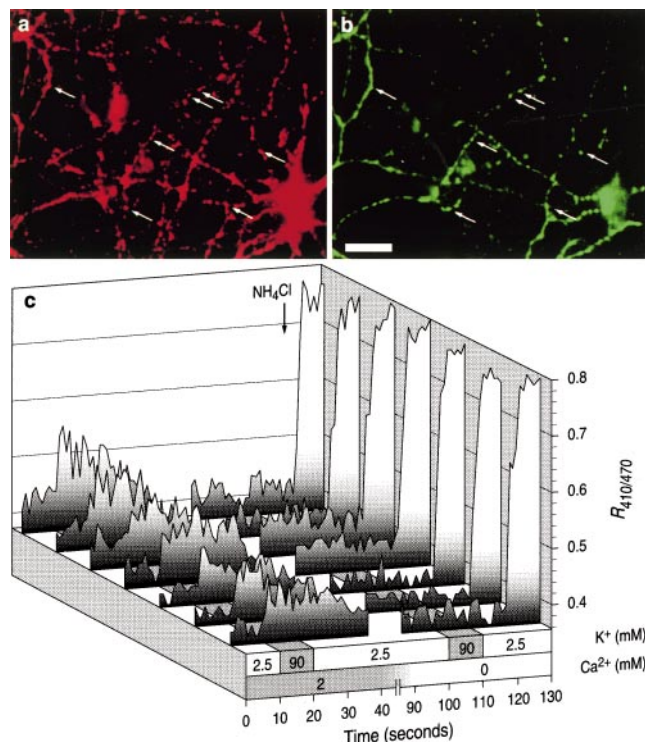


Figure 3 Neurotransmission visualized with ratiometric pHluorin. **a**, Hippocampal neurons immunostained with a monoclonal antibody against synaptotagmin-I. **b**, Synapses expressing synapto-pHluorin, formed by HSV-infected neurons whose somata lie outside the field of view. Arrows mark some points of registration between **a** and **b**. Scale bar, 20 μm . **c**, Recordings of synaptic activity from seven boutons, sampled at 1 Hz. Neurons were stepped through two depolarization cycles (by raising extracellular $[\text{K}^+]$ to 90 mM), the first in 2 mM extracellular Ca^{2+} and the second in the absence of Ca^{2+} and the presence of 2.5 mM EGTA.

attached to the inner surface of the vesicle membrane, will reflect this pH change in its excitation spectrum. Figure 3a shows a field of hippocampal neurons² forming abundant synapses revealed by immunostaining for synaptotagmin-I (ref. 6). Some of these synapses were made by neurons infected with a herpes simplex virus (HSV) vector^{2,19,20} expressing synapto-pHluorin, a fusion protein of VAMP-2/synaptobrevin⁶ and the ratiometric pHluorin, joined to VAMP at its lumenally exposed C terminus (Fig. 3b).

Under resting conditions, the synapto-pHluorin reported an intravesicular pH (mean $\pm$ s.d.) of 5.67 ± 0.71 ($n = 84$). Depolarization with 90 mM KCl elicited a prompt increase in $R_{410/470}$ which could be recorded from many boutons in parallel (Fig. 3c). The response depended on external Ca^{2+} (Fig. 3c and Table 1) and was abolished by a 24-hour preincubation with 10 nM botulinum neurotoxin B (BoNT; ref. 6) (Table 1). $R_{410/470}$ peaked at 8–20% of the signal that would have resulted from simultaneous fusion of all of the vesicles in the bouton; the latter was estimated by neutralizing the pH in all vesicles with 50 mM NH_4Cl (Fig. 3c). The increase in $R_{410/470}$ most probably corresponds to the 'readily releasable' pool^{6,21} of one to two dozen²¹ of the 100–200 synaptic vesicles at an active zone⁶.

Following repolarization, $R_{410/470}$ gradually declined (Fig. 3c). Vesicle membrane protein is re-internalized by endocytosis⁶ to regenerate synaptic vesicles, and synapto-pHluorin is expected to return to its spectral baseline as the vesicle acidifies. The $R_{410/470}$ value, whether it pertains to a single synapse or to a population of many synapses in a region, therefore provides a running average of activity during the previous several seconds.

Ecliptic pHluorins, because they are non-fluorescent at pH < 6 under 470-nm excitation (Fig. 1c), eliminate background due to resting vesicles and are thus potentially suited for detecting single vesicle fusion events. To test this in a favourable case, ecliptic synapto-pHluorin (a fusion protein of VAMP and the ecliptic pHluorin) was expressed in the mast cell line RBL-2H3 (ref. 22). The protein localized to scattered fluorescent puncta, which under resting conditions were seen only with 410- but not 470-nm excitation (Fig. 4a, frame 1); the pH in these secretory granules, measured with the ratiometric pHluorin, averaged (mean $\pm$ s.d.) 5.20 ± 0.55 ($n = 29$).

After initiating a secretory response²², granule content was released into the medium (Fig. 4b), and changes in fluorescence excited at 470 nm occurred in locations harbouring granules (Fig. 4a, frames 1–7). Individual spots of variable integrated intensity (Fig. 4c) appeared suddenly at various times after the stimulus to secrete (Fig. 4a, frames 2–7). The appearance of fluorescence followed the

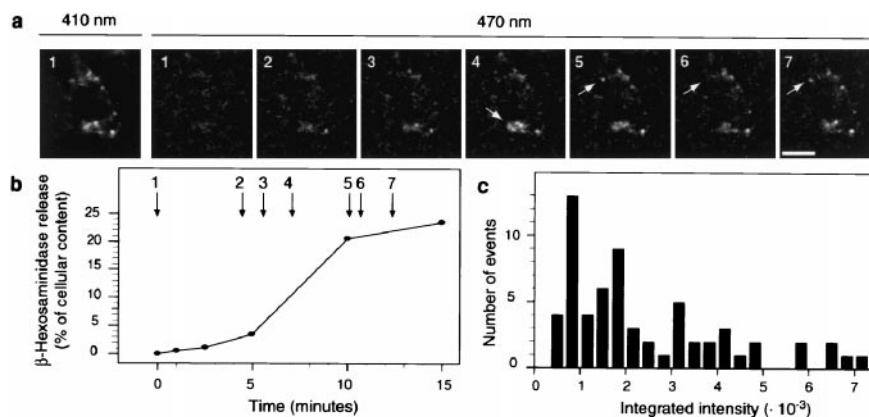


Figure 4 Secretion visualized with ecliptic pHluorin. **a**, RBL-2H3 cells under resting conditions (frame 1) and after triggering exocytosis (frames 2–7). Arrows mark a cluster from which individual spots are disappearing (frame 4), and a 'flickering' spot (frames 5–7). Images were contrast-enhanced and low-pass-filtered with a 3×3 binomial kernel. Scale bar, 10 μm . **b**, Release of β -hexosami-

nidase, a marker enzyme of RBL cell granules²². Times at which frames in **a** were acquired are indicated by arrows. **c**, Size distribution of secretory events. Integrated intensity is the product of pixel area and average fluorescence intensity.

appearance of granule content in the medium (Fig. 4b), as would be expected if each event were due to fusion of a single granule²³, or of multiple granules undergoing compound exocytosis^{23,24}.

Occasionally, fluorescent spots disappeared, indicating granule retrieval and re-acidification (follow the cluster marked by the arrow in frame 4 through frames 5 and 6). Within the time frame of our experiments (Fig. 4b), these 'off' events²³ were less frequent than the 'on' events of exocytosis, consistent with slow resolution of the often tortuous membrane topology created at exocytosis sites²⁴. Rarely, a granule appeared, disappeared, and reappeared in what seemed to be the same location (arrows in frames 5–7). This could correspond to the phenomenon of 'flicker'²³, in which transient opening and closing of a fusion pore would cause the vesicle's internal pH (and the pHluorin's emission intensity) to fluctuate.

The pHluorin technique may be extended to preparations other than cultured cells, trafficking processes other than exocytosis, and read-out formats other than fluorescence microscopy. For examples, ecliptic synapto-pHluorins imaged by two-photon or confocal microscopy could sensitively report transmission in neural tissues of transgenic animals, in a transmitter- or cell type-specific manner. Generally, pHluorins should be capable of monitoring pathways that connect compartments of differing pH, from receptor-mediated endocytosis⁵ to internalization of activated signalling receptors^{25,26} or insulin-dependent translocation of glucose transporters from internal stores to the cell surface²⁷. Many of these processes involve potential drug targets, and pHluorins could be adapted to cell- or tissue-based assays for high-throughput screening of compound libraries. □

Methods

Mutagenesis. Wild-type *Aequorea victoria* gfp cDNA (pGFP-1; Clontech) was subjected to PCR mutagenesis. Directed codon changes were introduced with *Pwo* polymerase; combinatorial steps (codons 94–97, 146–149, 164–168, 202–205, 221–225) used primer libraries of 32- to 32,768-fold nucleotide degeneracy and *Taq* polymerase. PCR products were ligated to pGEMEX-2 (Promega), transformed into *E. coli*, and visibly fluorescent colonies expanded for analysis. Clones were grown, lysed (50 mM Tris, pH 8.0, 2 mM EDTA, 0.2 mg ml⁻¹ lysozyme, 200 U ml⁻¹ DNase I), and analysed in 96-well plates. Three fluorescence readings at 510 nm were obtained in a Labsystems Fluoroskan II plate reader equipped with 400- and 460-nm excitation filters: the first at pH 8.0, the second after addition of acid to reduce the pH to 6.0, and the third after addition of base to revert the pH to 7.4. The plasmid encoding the mutant with the largest reversible change served as the PCR template in the next round.

cDNAs encoding ratiometric and ecliptic pHluorins—both of which also carry two amino-acid changes (V163A, S175G) that are known to improve folding²⁸ at 37°C—were sequenced and expressed in pGEX-2T (Pharmacia). Fluorescence spectra were recorded on material obtained after thrombin cleavage of the respective glutathione-S-transferase (GST) fusion protein, in a Perkin-Elmer LS-50B spectrofluorimeter with 8-nm bandwidths for excitation and emission. Chromophore concentrations are based on the 447-nm absorbance at pH 11.8 of denatured GFP or pHluorin and an extinction coefficient⁷ of 44,100 M⁻¹ cm⁻¹. Response times to pH jumps were estimated with the help of an SFA-20 rapid kinetics accessory (Hi-Tech Scientific).

Cells. GFP modules were linked to targeting modules via two -Ser-Gly-Gly-repeats and one -Thr-Gly-Gly- repeat. The insertion sites were placed between signals for translocation into the endoplasmic reticulum and GPI-anchor addition (derived from preprolactin and decay accelerating factor¹⁶, respectively), at the mature N terminus of TGN38 (ref. 17) and at the C termini of VAMP⁶ and cellubrevin¹⁸.

The vector pCI (Promega) was used to drive expression in HeLa and RBL-2H3 cells. Cell-surface IgE receptors of RBL-2H3 cells were loaded with 10 ng ml⁻¹ anti-DNP IgE (Sigma) and secretory responses triggered with 1 µg ml⁻¹ anti-IgE antibodies (PharMingen). β-Hexosaminidase was assayed as described²².

Hippocampal neurons were infected with THZ.3 (α4⁻, α22⁻, α27⁻, U_L 41⁻)

HSV virions¹⁹ carrying amplicon plasmids based on the pα4'a' backbone²⁰. Neural cultures were prepared as described², except that hippocampi were collected from E19 rats, horse serum was used, and 10 µM cytosine arabinoside was added from day 3 after plating. Experiments were done after 17–29 days *in vitro*.

Microscopy. Two days after transfection or viral infection, cells in imaging buffer (25 mM Na-HEPES, pH 7.4, 119 mM NaCl, 2.5 mM KCl, 2 mM CaCl₂, 2 mM MgCl₂, 30 mM glucose) were analysed at 37°C on a Zeiss Axiovert microscope equipped with a ×40, 1.3 NA Plan-Neofluar objective and ×1.6 and ×2.5 Optovar inserts. Experiments were controlled through MetaFluor 3.0 (Universal Imaging), with off-line background subtraction and image analysis, using MetaMorph 3.0 (Universal Imaging) and Mathematica 3.0 (Wolfram Research). A Polychrome II monochromator (Till Photonics) delivered 12-nm excitation bands centred at 410 and 470 nm and found optimal for imaging. Emitted light passed through a dichromatic mirror (500DCXR) and a bandpass filter (HQ535/50, both from Chroma Technologies) and was collected on an intensified PentaMAX-512EFT frame-transfer camera (Princeton Instruments).

Received 30 January; accepted 1 May 1998.

1. Tsien, R. Y. Fluorescent probes of cell signaling. *Annu. Rev. Neurosci.* **12**, 227–253 (1989).
2. Miesenböck, G. & Rothman, J. E. Patterns of synaptic activity in neural networks recorded by light emission from synaptotagmins. *Proc. Natl Acad. Sci. USA* **94**, 3402–3407 (1997).
3. Miyawaki, A. et al. Fluorescent indicators for Ca²⁺ based on green fluorescent proteins and calmodulin. *Nature* **388**, 882–887 (1997).
4. Romoser, V. A., Hinkle, P. M. & Persechini, A. Detection in living cells of Ca²⁺-dependent changes in the fluorescence emission of an indicator composed of two green fluorescent protein variants linked by a calmodulin-binding sequence. *J. Biol. Chem.* **272**, 13270–13274 (1997).
5. Anderson, R. G. & Orci, L. A view of acidic intracellular compartments. *J. Cell Biol.* **106**, 539–543 (1988).
6. Südhof, T. C. The synaptic vesicle cycle: a cascade of protein-protein interactions. *Nature* **375**, 645–653 (1995).
7. Ward, W. W. in *Bioluminescence and Chemiluminescence* (eds DeLuca, M. A. & McElroy, W. D.) 235–242 (Academic, New York, 1981).
8. Ward, W. W., Prentice, H. J., Roth, A. F., Cody, C. W. & Reeves, S. C. Spectral perturbations of the *Aequorea* green-fluorescent protein. *Photochem. Photobiol.* **35**, 803–808 (1982).
9. Heim, R., Prasher, D. C. & Tsien, R. Y. Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc. Natl Acad. Sci. USA* **91**, 12501–12504 (1994).
10. Chatteraj, M., King, B. A., Bublitz, G. U. & Boxer, S. G. Ultra-fast excited state dynamics in green fluorescent protein: Multiple states and proton transfer. *Proc. Natl Acad. Sci. USA* **93**, 8362–8367 (1996).
11. Ormö, M. et al. Crystal structure of the *Aequorea victoria* green fluorescent protein. *Science* **273**, 1392–1395 (1996).
12. Yang, F., Moss, L. G. & Phillips, G. N. Jr The molecular structure of green fluorescent protein. *Nature Biotechnol.* **14**, 1246–1251 (1996).
13. Brejc, K. et al. Structural basis for dual excitation and photoisomerization of the *Aequorea victoria* green fluorescent protein. *Proc. Natl Acad. Sci. USA* **94**, 2306–2311 (1997).
14. Ehrig, T., O'Kane, D. J. & Prendergast, E. G. Green-fluorescent protein mutants with altered fluorescence excitation spectra. *FEBS Lett.* **367**, 163–166 (1995).
15. Heim, R. & Tsien, R. Y. Engineering green fluorescent protein for improved brightness, longer wavelengths and fluorescence resonance energy transfer. *Curr. Biol.* **6**, 178–182 (1996).
16. Caras, I. W., Weddell, G. N., Davitz, M. A., Nussenzweig, V. & Martin, D. W. Jr Signal for attachment of a phospholipid membrane anchor in decay accelerating factor. *Science* **238**, 1280–1283 (1987).
17. Luzio, J. P. et al. Identification, sequencing and expression of an integral membrane protein of the trans-Golgi network (TGN38). *Biochem. J.* **270**, 97–102 (1990).
18. McMahon, H. T. et al. Cellubrevin is a ubiquitous tetanus-toxin substrate homologous to a putative synaptic vesicle fusion protein. *Nature* **364**, 346–349 (1993).
19. Marconi, P. et al. Replication-defective herpes simplex virus vectors for gene transfer *in vivo*. *Proc. Natl Acad. Sci. USA* **93**, 11319–11320 (1996).
20. Lawrence, M. S., Ho, D. Y., Dash, R. & Sapolsky, R. M. Herpes simplex virus vectors overexpressing the glucose transporter gene protect against seizure-induced neuron loss. *Proc. Natl Acad. Sci. USA* **92**, 7247–7251 (1995).
21. Stevens, C. F. & Tsujimoto, T. Estimates for the pool size of releasable quanta at a single central synapse and for the time required to refill the pool. *Proc. Natl Acad. Sci. USA* **92**, 846–849 (1995).
22. Roa, M., Paumet, F., Le Mao, J., David, B. & Blank, U. Involvement of the *ras*-like GTPase rab3d in RBL-2H3 mast cell exocytosis following stimulation via high-affinity IgE receptors (FcεRI). *J. Immunol.* **159**, 2815–2823 (1997).
23. Fernandez, J. M., Neher, E. & Gomperts, B. D. Capacitance measurements reveal stepwise fusion events in degranulating mast cells. *Nature* **312**, 453–455 (1984).
24. Chandler, D. E. & Heuser, J. E. Arrest of membrane fusion events in mast cells by quick-freezing. *J. Cell Biol.* **96**, 666–674 (1980).
25. Ullrich, A. & Schlessinger, J. Signal transduction by receptors with tyrosine kinase activity. *Cell* **61**, 203–212 (1990).
26. Yu, S. S., Lefkowitz, R. J. & Hausdorff, W. P. β-Adrenergic receptor sequestration: A potential mechanism of receptor resensitization. *J. Biol. Chem.* **268**, 337–341 (1993).
27. James, D. E. & Piper, R. C. Insulin resistance, diabetes, and the insulin-regulated trafficking of GLUT-4. *J. Cell Biol.* **126**, 1123–1126 (1994).
28. Siemering, K. R., Golbik, R., Sever, R. & Haseloff, J. Mutations that suppress the thermosensitivity of green fluorescent protein. *Curr. Biol.* **6**, 1653–1663 (1996).

Acknowledgements. We thank R. Miller for technical assistance, D. Krisky and J. Glorioso for HSV strain THZ.3 and 7B cells, G. Schiavo for BoNT, the Fonds de la Recherche en Santé du Québec for a postdoctoral fellowship (to D.A.D.), and Q. Al-Awqati for discussion. This research was supported by the G. Harold and Leila Y. Mathers Charitable Foundation.

Correspondence and requests for materials should be addressed to J.E.R. The GenBank accession numbers for ratiometric and ecliptic pHluorin are AF058694 and AF058695, respectively.

Antagonism between *extradenticle* function and Hedgehog signalling in the developing limb

Sergio González-Crespo^{*†‡}, Muna Abu-Shaar^{‡||}, Miguel Torres[§], Carlos Martínez-A[§], Richard S. Mann[‡] & Ginés Morata^{*}

^{*} Centro de Biología Molecular, CSIC-UAM, Universidad Autónoma de Madrid, 28049 Madrid, Spain

[‡] Department of Biochemistry and Molecular Biophysics, and Integrated Program in Cellular, Molecular & Biophysical Studies, Columbia University College of Physicians and Surgeons, 701 West 168th Street, New York, New York 10032, USA

[§] Departamento de Inmunología y Oncología, Centro Nacional de Biotecnología, Universidad Autónoma de Madrid, 28049 Madrid, Spain

^{||} These authors contributed equally to this work

[†] Present address: Centre d'Investigació i Desenvolupament, CSIC, Jordi Girona 18-26 08034 Barcelona, Spain

The *Drosophila* homeobox gene *extradenticle* (*exd*) encodes a highly conserved cofactor of Hox proteins^{1–3}. *exd* activity is regulated post-translationally by a mechanism involving nuclear translocation^{4,5}; only nuclear Exd protein is functional. The *exd* gene is required for patterning of the proximal region of the leg^{6,7}, whereas patterning of the distal region requires signalling by the

Wingless (*Wg*) and Decapentaplegic (*Dpp*)^{8,9} proteins, which are in turn activated by Hedgehog (*Hh*)¹⁰. Here we show that *exd* function and *Dpp*/*Wg* signalling are antagonistic and divide the leg into two mutually exclusive domains. In the proximal domain, *exd* activity prevents cells from responding to *Dpp* and *Wg*. Conversely, in the distal domain, *exd* function is suppressed by the *Dpp*/*Wg* response gene *Distal-less* (*Dll*), which prevents the nuclear transport of Exd. We also found that the product of a murine homologue of *exd* (*Pbx1*) is regulated at the subcellular level, and that its pattern of nuclear localization in the mouse limb resembles that of Exd in the *Drosophila* leg. These findings suggest that the division of the limb into two antagonistic domains, as defined by *exd* (*Pbx1*) function and *Hh* signalling, may be a general feature of limb development.

The *Drosophila* imaginal leg disc contains a proximal region where Exd is nuclear (Exd-nuc) and a distal region (Exd-cyt) where it is cytoplasmic^{4,5}. Only the nuclear form is functional and is involved in patterning in the proximal leg^{6,7}. The development of the Exd-cyt domain requires the activity of *Dll*¹¹, a gene activated in the leg primordium during early embryogenesis at the intersection of the stripes of cells expressing *dpp* and *wg*¹². When *Dll* is first expressed, Exd is nuclear (Fig. 1a–c), but by late embryogenesis Exd becomes cytoplasmic in *Dll*-expressing cells (Fig. 1d–f). At this stage the leg primordium also includes a ring of cells expressing the gene *escargot* (*esg*), but not *Dll*¹³; these cells have nuclear Exd (Fig. 1g, h). This subdivision into Exd-nuc and Exd-cyt domains persists for the rest of leg disc development.

We have found that the Exd-cyt region corresponds almost exactly with the sum of the expression domains of *Dll* and

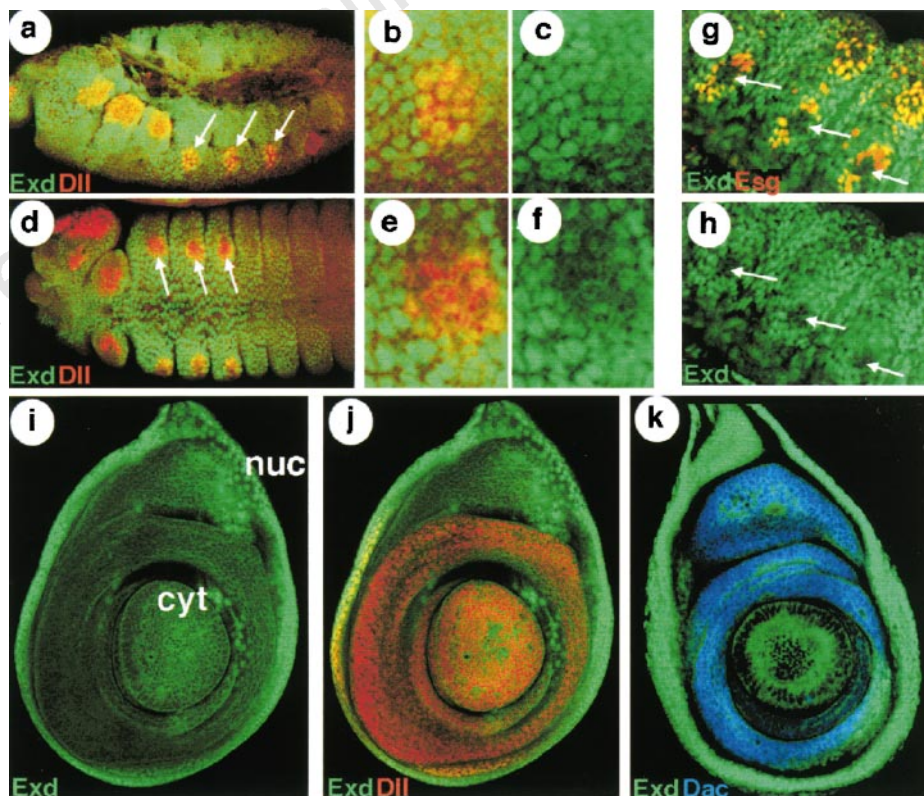


Figure 1 Subdivision of the leg primordium and the leg disc into two distinct domains. **a–c**, Stage 11 embryo stained for Exd (green) and *Dll* (red). **a**, Arrows point to the three leg primordia, one of which (T1) is magnified in **b** and **c** to show the nuclear co-localization of Exd and *Dll*. **d–f**, Stage 14 embryo stained as in **a–c**. **d**, Arrows point to the three thoracic primordia, one of which (T1) is magnified in **e** and **f** to show that Exd is not nuclear in cells expressing *Dll*. **g, h**, Thoracic embryonic segments stained with Exd (green) and *Esg* (red). Arrows point to the leg primordia that include cells expressing *Dll* and *esg*¹³. Note that the cells

stained with *Esg* contain nuclear Exd. **i–k**, Mature leg discs stained for Exd, *Dll* and *Dac* proteins. **i**, The central region of cytoplasmic (cyt) localization and the peripheral region of nuclear (nuc) localization of the Exd protein; **j** and **k** also show the expression of *Dll* (red) and *dac* (blue), respectively. There is a small region (yellow) of overlap in **j** corresponding to a late tier of *Wg*- and *Dpp*-independent *Dll* expression in the proximal region⁸. The sum of the *Dll* and *dac* domains corresponds to the entire region in which Exd is cytoplasmic.

dachshund (*dac*)¹⁴, two genes activated in response to Wg and Dpp⁹ (Fig. 1i–k). The expression of a third Dpp response gene, *optomotor-blind* (*omb*)¹⁵, is also confined within the Exd-cyt domain and abuts the boundary between Exd-nuc and Exd-cyt¹⁶. This suggests that the Hh pathway operates only in the Exd-cyt domain, and is consistent with previous studies^{8,16} showing that reduction of *hh* function results in loss of distal structures but has no effect on proximal structures where Exd is nuclear.

Strikingly, mice mutant for *Sonic hedgehog* (*Shh*) present a limb phenotype resembling that of *hh*[−] in *Drosophila*: the distal regions

are lacking but the proximal regions are unaltered¹⁷. As this may suggest comparable organization of the limb, we examined the expression of the *Pbx* genes¹⁸ (the vertebrate homologues of *exd*). We found that the subcellular distribution of Pbx1 resembles that of Exd in *Drosophila*: in the proximal mesenchyme the Pbx1 protein is in the nuclei (Fig. 2a, c, e), whereas in the distal mesenchyme it can be detected only in the cytoplasm (Fig. 2a, b, d). Thus, the *Drosophila* leg and the mouse limb are subdivided into proximal and distal domains as defined by the subcellular localization of Exd (or Pbx1), and in both cases *hh* function appears to be similarly

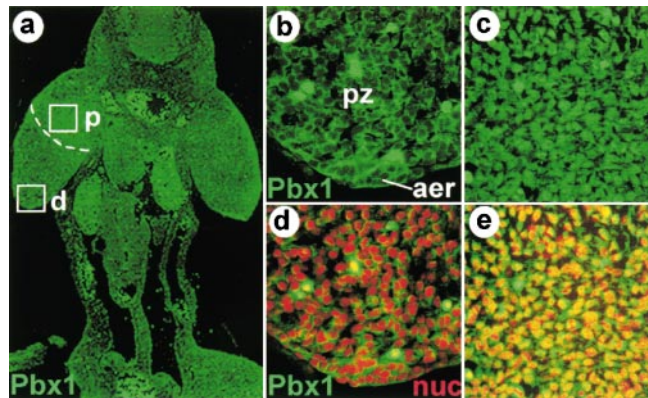


Figure 2 *Pbx1* expression in the mouse limb primordium. **a**, Low magnification of the anterior body regions of a 10.5 d.p.c. (days post-coitum) mouse embryo stained with anti-Pbx1 antibody (green). The discontinuous line across the limb marks the boundary between nuclear and cytoplasmic staining of Pbx1. The two squares mark proximal (p) and distal (d) zones of the left limb, which are amplified in **b–e**. **b**, **d**, High magnification of the distal zone (d in **a**) to show cytoplasmic

localization of Pbx1 protein. Hoetsch staining of nuclei (nuc) appears red. **c**, **e**, High magnification of the proximal zone (p in **a**) to show the nuclear localization of Pbx1 in those cells. The yellow in the nuclei of **e** corresponds to the combination of Exd (green) and Hoetsch (red) staining. pz, progress zone; aer, apical ectodermal ridge.

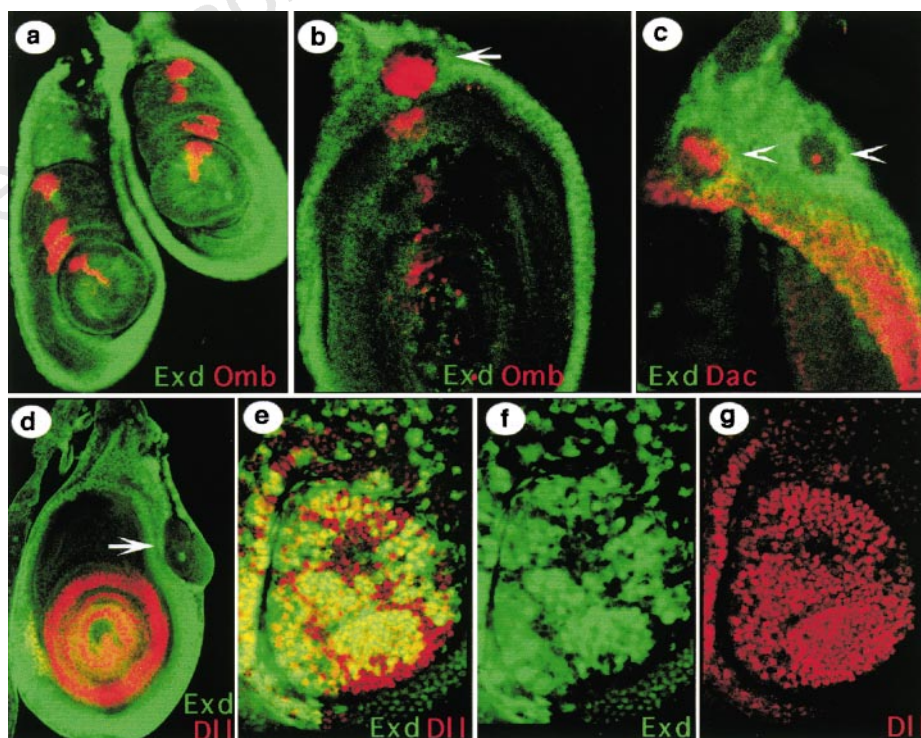


Figure 3 Differential effects of *exd* function on *omb*, *dac* and *Dll* expression in the leg disc. **a**, Wild-type expression of *exd* (green) and *omb* (red). Note that *omb* is expressed only in the dorsal part of the Exd cytoplasmic domain and that its proximal limit abuts the zone of nuclear Exd. **b**, Dorsal *exd*[−] clone (arrow) within the *exd* nuclear domain showing ectopic *omb* expression. **c**, Two *exd*[−] clones (arrowheads) in the dorsal region of the *exd* nuclear domain showing ectopic *dac* expression (red); note that not all the cells within the clones contain Dac protein.

d, Lateral *exd*[−] clone (arrow) showing no ectopic *Dll* expression (red). **e**, The central portion of a leg disc of a *Dll*-Gal4/UAS-*exd* larva stained for Exd (green) and *Dll* (red) showing coincidental localization of the two proteins in many nuclei (yellow). **f**, The same region of the disc showing the Exd protein present in many nuclei of the *Dll* domain. **g**, The localization of *Dll* protein is not altered by the presence of nuclear Exd.

required in the distal domain. We have studied in *Drosophila* the functional interactions between these domains by modifying either *exd* activity or Hh (Wg/Dpp) signalling.

Figure 3 shows the effects of altering *exd* function on the expression of three genes, *omb*, *dac* and *Dll*, which are known to respond to Dpp and/or Wg signals^{9,15,19}. The *omb* gene is normally expressed in the dorsal anterior region, coincident with Dpp (Fig. 3a). We found that *exd*⁻ clones in the dorsal region of the Exd-nuc

domain show ectopic expression of *omb* (Fig. 3b), whereas forcing nuclear Exd in the Exd-cyt domain eliminates *omb* expression (data not shown). Our results with *dac* were similar: *exd*⁻ clones often show ectopic *dac* expression (Fig. 3c), although not all the *exd*⁻ cells show *dac* activity. This indicates that, although *dac* activation requires elimination of *exd*, other factors are involved, possibly levels of Dpp and Wg⁹. In contrast to the effects on *omb* and *dac*, *exd*⁻ clones in the proximal region do not show activation of *Dll* (Fig. 3d).

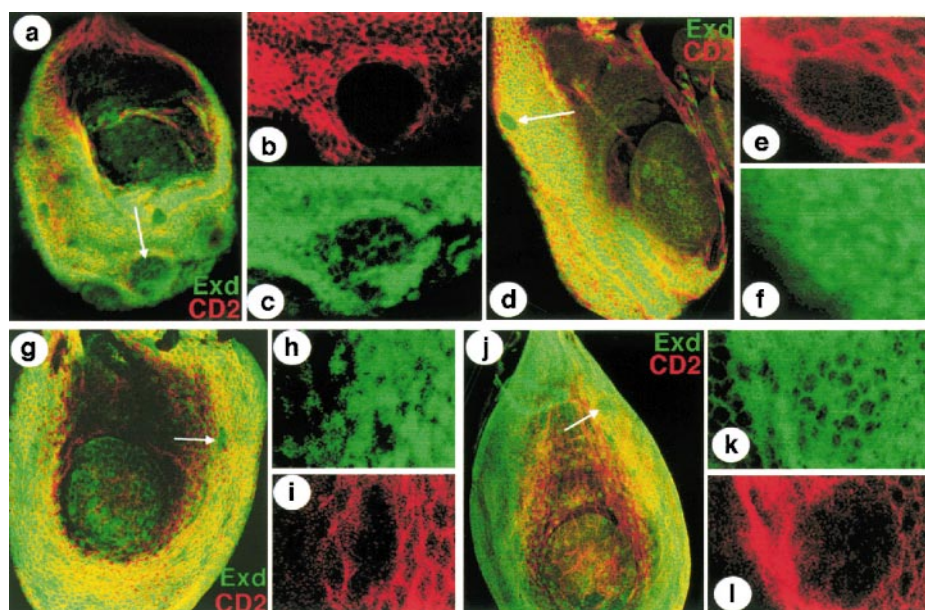


Figure 4 Dpp and Wg signalling and Exd subcellular localization. **a-f**, The effect of *tkv*^{OD} clones. **a**, Mature leg disc, stained for Exd (green) and CD2 (red), containing several *tkv*^{OD} clones labelled by the loss of CD2. A ventral clone (arrow) is magnified in **b** and **c** to show the cytoplasmic localization of Exd. In contrast, the

dorsal-lateral clone in **d** (arrow), **e** and **f** retains nuclear Exd localization. **g-i**, The effect of *wg*-expressing clones marked by the absence of CD2. **g**, A lateral clone (arrow), magnified in **h** and **i**, retaining nuclear localization of Exd. **j**, A dorsal clone (arrow) showing (magnified in **k** and **l**) cytoplasmic localization of Exd.

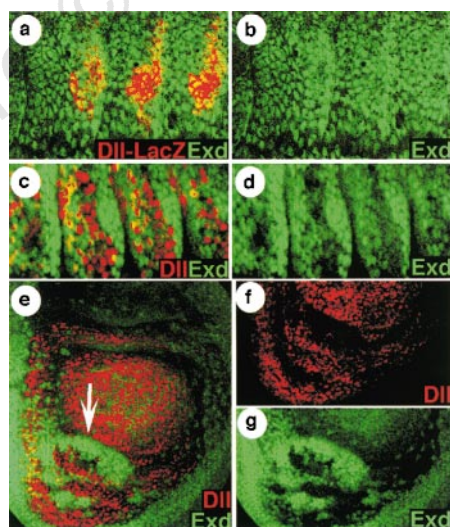
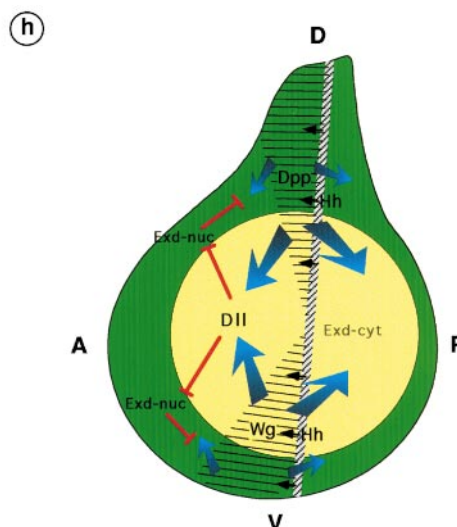


Figure 5 Control of Exd subcellular localization by *Dll*. **a, b**, The three thoracic segments of an embryo homozygous for a *Dll*^{SAT} mutant stained for β -gal and Exd. These embryos also contained the *Dll* enhancer-*lacZ* transgene, *Dll*[304]-*lacZ* (ref. 29), which allowed identification of the leg primordia by staining for β -Gal. The leg primordia appear yellow in **a** because both Exd and β -gal are nuclear; in **b** only the Exd stain is shown. **c, d**, Thoracic segments of an embryo of genotype *ptc*-Gal4/UAS-*Dll* stained for *Dll* (red) and Exd (green). The ectopic expression of *Dll* in the anterior compartments results in cytoplasmic Exd protein. **e-g**, A *Dll*⁻ clone (arrow) in the *Dll* domain showing high levels of nuclear Exd. **h**, Model of the subdivision of the leg imaginal disc into two antagonistic domains, as defined by



the Hh pathway and *exd* function. The short-range inducer Hh (short black arrows) acts across the anteroposterior (A/P) compartment border (shadowed bar) to induce *dpp* and *wg* in the dorsal (D) and ventral (V) sides of the disc, respectively. In the Exd-cyt (yellow) domain, the concentrations of Dpp and Wg products (blue arrows in gradient) are sufficient to induce *Dll* activity, which prevents (red lines) nuclear transport of Exd and hence its function. In the Exd-nuc domain (green), Dpp and Wg are not able to induce *Dll* and Exd is transported to the nuclei, probably through the function of *hth*²¹. The activity of *exd* does not prevent the diffusion of the Dpp and Wg signals but blocks (red lines) the response of genes like *omb* and *dac*.

Moreover, forcing *exd* nuclear expression in the Exd-cyt domain does not result in the downregulation of *Dll* (Fig. 3e–g).

Because activation of *omb* and *dac* requires Dpp and Wg^{9,15,19}, these results indicate that the Dpp and Wg signals reach the Exd-nuc domain in sufficient concentrations to activate these and perhaps other genes, but that responses to these signals are blocked by Exd. In turn, the lack of *Dll* activation in *exd*[−] clones suggests that the Exd-nuc domain is outside the territory in which the Dpp and Wg signals can activate *Dll* (see below).

The finding that Exd blocks the response to Dpp and Wg led us to examine the consequences of forcing higher levels of these signals in the Exd-nuc domain. The experiments consisted of generating in the Exd-nuc domain clones of cells expressing a constitutively active form of the Dpp receptor *thick veins* (*tkv*^{QD})²⁰ or expressing a membrane-tethered form of Wg (see Methods). The effect of the *tkv*^{QD} clones depends on their position within the Exd-nuc domain. Dorsal clones, located close to the endogenous source of Dpp, do not alter the localization of Exd, whereas in ventral and ventral–lateral clones (close to the endogenous source of Wg) Exd becomes cytoplasmic (Fig. 4a–f). Ventral *tkv*^{QD} clones express *Dll*, as previously reported⁹. Similarly, clones of wg-expressing cells also prevent the nuclear translocation of Exd, but only in dorsal or dorsal–lateral clones (close to the endogenous source of Dpp) is Exd shifted to the cytoplasm; in ventral clones, close to the endogenous source of Wg, no effect is observed (Fig. 4g–l). These results indicate that Dpp/Wg signalling represses *exd* activity by preventing the nuclear transport of Exd protein. The topological restriction of the effect to ventral (in *tkv*^{QD}) or dorsal (in wg-expressing) clones indicates that it requires input from both Dpp and Wg signals. Further, these data demonstrate that both signals are normally present in the Exd-nuc domain.

Because *Dll* is activated in clones expressing *tkv*^{QD} or wg (ref. 9), we examined the possibility that it may be involved in regulating the subcellular localization of Exd. We found that, during both embryogenesis (Fig. 5a–d) and imaginal development (Fig. 5a–g), *Dll* activity prevents the nuclear translocation of Exd. However, alterations of *Dll* activity only affect Exd localization if induced before 72 h of development, suggesting that there may be other regulators. We considered *dac* as a candidate because it is expressed in a zone of the Exd-cyt domain where *Dll* is not expressed (Fig. 1k), but we failed to find any effect on the localization of Exd in cells lacking *dac* activity (data not shown). Recently, the gene *homothorax* (*hth*)²¹ has been identified as a positive regulator of Exd nuclear translocation, and there is evidence (M.A.-S. and R.S.M., unpublished data) that *Dll* represses *hth*. It is therefore likely that the role of *Dll*, and perhaps of other regulators, is mediated by repression of *hth*.

Our results are summarized in Fig. 5h. We propose that the leg subdivision into Exd-nuc and Exd-cyt domains is the result of Hh signalling and its mutual antagonism with *exd* function. In the centre of the disc the Hh-response signals Dpp and Wg are in sufficient concentrations to induce *Dll*, which acts as a repressor of Exd nuclear translocation. Blocking the nuclear transport of Exd allows full expression of the Hh pathway in the Exd-cyt domain: *omb*, *dac* and probably other response genes as well are activated. In the peripheral region the concentrations of Dpp and Wg are too low to induce *Dll*, and Exd remains nuclear. Although Dpp and Wg could still induce *omb*, *dac* and perhaps other genes, *exd* activity blocks their activation.

The subdivision of the limb into proximal and distal antagonistic domains is analogous to the subdivision of the leg along the dorsoventral axis^{19,22,23}. In that case, two domains are generated because the Dpp and Wg signals are themselves mutually antagonistic^{19,22,23}. Thus, during development the leg disc is subdivided along both the dorsoventral and the proximodistal axes into discrete domains by mutual antagonism.

We have found that the subcellular distribution of mouse Pbx1 resembles that of Exd in *Drosophila*. This finding, together with the

remarkable similarity between the *Drosophila* and mouse limbs when the Hh pathway is inactivated^{18,16,17}, suggests a similar organization into two regions along the proximodistal axis, as defined by interactions between the *exd/Pbx* genes and the Hh pathway. □

Methods

Fly stocks, crosses and production of marked clones. We generated *exd*[−] clones by the FLP/FRT method in larvae heterozygous for *exd*^{XP11}, a null allele². Loss of *Dll* function in embryos and imaginal discs was analysed using the *Dll*^{SA1} and *Dll*^{MP} alleles^{9,12,24}. *Dll*[−] clones in the imaginal discs were made using a combination of the FRT/FLP and Minute²⁵ methods. Gain of *Dll* function was analysed in embryos from the cross *ptc*-GAL4 × UAS-*Dll* at 29 °C. Clones expressing the constitutively active form of the Dpp receptor (*tkv*^{QD}) or a membrane-tethered form of Wg were induced using the GAL4 flip-out method^{20,26}. Larvae of the genotype *hs-FLP/+; tsh-GAL4/+; UAS > CD2 > tkv*^{QD}/+ or *hs-FLP; tsh-GAL4/UAS > CD2y⁺ > Nrt-flu-wg* were heat shocked at 24–48 h before dissection and imaginal discs were stained using appropriate antibodies. Because the Gal4 protein is produced only in the *tsh* domain, which is coincidental with the *exd* functional domain¹⁶, the expression of *tkv*^{QD} or *Nrt-flu-wg* can occur only within the Exd-nuc domain.

Immunostains. Embryos or imaginal discs were stained by standard procedures for confocal microscopy^{4,5,16}. Specific antibodies used were rat or rabbit anti-Exd¹⁶, monoclonal anti-Dll¹², mouse anti-Dac¹⁴, rabbit anti-β-gal (Cappel) and mouse anti-CD2 (Serotec). For *esg* and *omb* expression we used the corresponding *lac-Z* lines^{15,27}. In the mouse limb, Pbx1 was detected as described²⁸, using a polyclonal antibody against the amino acids 20–39 of human Pbx1 (Santa Cruz). Optical sections were obtained using an Ar-Kr laser and either a TCS-NT Leica, Biorad MRC600 or Biorad MRC1024 confocal imaging systems.

Received 18 February; accepted 11 May 1998.

- Peifer, M. & Wieschaus, E. Mutations in the *Drosophila* gene *extradenticle* affect the way specific homeo domain proteins regulate segmental identity. *Genes Dev.* **4**, 1209–1223 (1990).
- Rauskolb, C., Peifer, M. & Wieschaus, E. *extradenticle*, a regulator of homeotic gene activity, is a homolog of the homeobox-containing human proto-oncogene *pbx-1*. *Cell* **74**, 1101–1112 (1993).
- Mann, R. & Chan, S. K. Extra specificity from *extradenticle*: the partnership between Hox and *exd/Pbx* homeodomain proteins. *Trends Genet.* **12**, 258–262 (1996).
- Mann, R. & Abu-Shaar, M. Nuclear import of the homeodomain protein *Extradenticle* in response to Wg and Dpp signalling. *Nature* **383**, 630–633 (1996).
- Aspland, S. E. & White, R. A. H. Nucleocytoplasmic localization of *extradenticle* protein is spatially regulated throughout development in *Drosophila*. *Development* **124**, 741–747 (1997).
- González-Crespo, S. & Morata, G. Control of *Drosophila* adult pattern by *extradenticle*. *Development* **121**, 2117–2125 (1995).
- Rauskolb, C., Smith, K. M., Peifer, M. & Wieschaus, E. *extradenticle* determines segmental identities throughout *Drosophila* development. *Development* **121**, 3663–3673 (1995).
- Diaz-Benjumea, F. J., Cohen, B. & Cohen, S. M. Cell interaction between compartments establishes the proximal–distal axis of *Drosophila* legs. *Nature* **372**, 175–179 (1994).
- Lecuit, T. & Cohen, S. M. Proximal–distal axis formation in the *Drosophila* leg. *Nature* **388**, 139–145 (1997).
- Basler, K. & Struhl, G. Compartment boundaries and the control of *Drosophila* limb pattern by *hedgehog* protein. *Nature* **368**, 208–214 (1994).
- Cohen, S. M. & Jürgens, G. Proximo-distal pattern formation in *Drosophila*: cell autonomous requirements for *Distal-less* gene activity in limb development. *EMBO J.* **8**, 2045–2055 (1989).
- Cohen, B., Simcox, A. A. & Cohen, S. M. Allocation of the thoracic imaginal primordia in the *Drosophila* embryo. *Development* **117**, 597–608 (1993).
- Goto, S. & Hayashi, S. Specification of the embryonic limb primordium by graded activity of *Decapentaplegic*. *Development* **124**, 125–132 (1997).
- Mardon, G., Solomon, N. M. & Rubin, G. M. *dachshund* encodes a nuclear protein required for normal eye and leg development in *Drosophila*. *Development* **120**, 3473–3486 (1994).
- Grim, S. & Pflugfelder, G. O. Control of the gene *optomotor-blind* in *Drosophila* wing development by *decapentaplegic* and *wingless*. *Science* **271**, 1601–1604 (1996).
- González-Crespo, S. & Morata, G. Genetic evidence for the subdivision of the arthropod limb into coxopodite and telopodite. *Development* **122**, 3921–3928 (1996).
- Chiang, C. et al. Cyclopia and defective axial patterning in mice lacking *Sonic hedgehog* gene function. *Nature* **383**, 407–413 (1996).
- Monica, K., Galili, N., Nourse, J., Saltman, D. & Cleary, M. L. PBX2 and PBX3, new homeobox genes with extensive homology of the human protooncogene PBX1. *Mol. Cell. Biol.* **11**, 6149–6157 (1991).
- Brook, W. J. & Cohen, S. M. Antagonistic interactions between wingless and decapentaplegic responsible for dorsal-ventral pattern in the *Drosophila* leg. *Science* **273**, 1373–1377 (1996).
- Nellen, D., Burke, R., Struhl, G. & Basler, K. Direct and long-range action of a Dpp morphogen gradient. *Cell* **85**, 357–368 (1996).
- Rieckhof, G., Casares, F., Ryoo, H. D., Abu-Shaar, M. & Mann, R. Nuclear translocation of *Extradenticle* requires *homothorax*, which encodes an extradenticle-related homeodomain protein. *Cell* **91**, 171–183 (1997).
- Penton, A. & Hoffman, F. M. *Decapentaplegic* restricts the domain of *wingless* during *Drosophila* limb patterning. *Nature* **382**, 162–165 (1996).
- Jiang, J. & Struhl, G. Complementary and mutually exclusive activities of *Decapentaplegic* and *Wingless* organize axial patterning during *Drosophila* leg development. *Cell* **86**, 401–409 (1996).
- Cohen, S. M., Brönnner, G., Künter, E., Jürgens, G. & Jäckle, H. *Distal-less* encodes a homeodomain protein required for limb development in *Drosophila*. *Nature* **338**, 432–434 (1989).
- Morata, G. & Ripoll, P. *Minutes*: Mutants of *Drosophila* autonomously affecting cell division rate. *Dev. Biol.* **42**, 211–221 (1975).

26. Zecca, M., Basler, K. & Struhl, G. Direct and long range action of a wingless morphogen gradient. *Cell* **87**, 833–844 (1996).
27. Fuse, N., Hirose, S. & Hayashi, S. Determination of wing cell fate by the *escargot* and *snail* genes in *Drosophila*. *Development* **122**, 1059–1067 (1996).
28. Torres, M., Gómez-Pardo, E., Dressler, G. R. & Gross, P. Pax2 controls multiple steps of urogenital development. *Development* **121**, 4057–4065 (1995).
29. Vachon, G. *et al.* Homeotic genes of the bithorax complex repress limb development in the abdomen of the *Drosophila* embryo through the target gene *Distal-less*. *Cell* **71**, 437–450 (1992).

Acknowledgements. We thank K. Basler, S. Carroll, S. Cohen, G. Mardon, G. Pflugfelder and G. Struhl for fly stocks and antibodies; F. Casares, E. Sánchez-Herrero and I. Guerrero for comments on the manuscript; J. Casanova for critical help; and A. Bosch for help with the confocal microscope. This work was supported by the Dirección General de Investigación Científica y Técnica (G.M.), a Human Frontiers grant (G.M. and R.S.M.), an institutional grant of the Fundación Ramón Areces to the Centro de Biología Molecular, Institutional grants of the CSIC and Pharmacia-Upjohn to the Departamento de Immunología y Oncología, and an NIH grant to R.S.M., who is a scholar of the Leukemia Society of America.

Correspondence and requests for materials should be addressed to G.M. (e-mail: gmorata@cbm.uam.es).

Reduction of atherosclerosis in mice by inhibition of CD40 signalling

François Mach^{*†}, Uwe Schönbeck^{*†}, Galina K. Sukhova^{*}, Elizabeth Atkinson^{*} & Peter Libby^{*}

^{*} Vascular Medicine and Atherosclerosis Unit, Cardiovascular Division, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, 221 Longwood Avenue, Boston, Massachusetts 02115, USA

[†] These authors contributed equally to this work

Increasing amounts of evidence support the involvement of inflammation and immunity in atherogenesis^{1–4}, but mediators of communication between the major cell types in atherosclerotic plaques are poorly defined. Cells in human atherosclerotic lesions express the immune mediator CD40 and its ligand CD40L (also

known as CD154 or gp39)⁵. The interaction of CD40 with CD40L figures prominently in both humoral and cell-mediated immune responses⁶. CD40L-positive T cells accumulate in atheroma⁵, and, by virtue of their early appearance, persistence and localization at sites of lesion growth and complication, activated T cells may coordinate important aspects of atherogenesis^{7–9}. Interruption of CD40L–CD40 signalling by administration of an anti-CD40L antibody limits experimental autoimmune diseases such as collagen-induced arthritis, lupus nephritis, acute or chronic graft-versus-host disease, multiple sclerosis and thyroiditis^{10–14}. Ligation of CD40 on atheroma-associated cells *in vitro* activates functions related to atherogenesis, including induction of pro-inflammatory cytokines⁵, matrix metalloproteinases^{15,16}, adhesion molecules^{17–19} and tissue factor^{16,20}. However, the role of CD40 signalling in atherogenesis *in vivo* remains unknown. Here we determine whether interruption of CD40 signalling influences atherogenesis *in vivo* in hyperlipidaemic mice. Treatment with antibody against mouse CD40L limited atherosclerosis in mice lacking the receptor for low-density lipoprotein that had been fed a high-cholesterol diet for 12 weeks. This antibody reduces the size of aortic atherosclerotic lesions by 59% and their lipid content by 79%. Furthermore, atheroma of mice treated with anti-CD40L antibody contained significantly fewer macrophages (64%) and T lymphocytes (70%), and exhibited decreased expression of vascular cell adhesion molecule-1. These data support the involvement of inflammatory pathways in atherosclerosis and indicate a role for CD40 signalling during atherogenesis in hyperlipidaemic mice.

We tested the hypothesis that administration of a neutralizing anti-CD40L monoclonal antibody could mitigate the development of atherosclerosis in mice. Male mice lacking the low-density lipoprotein (LDL) receptor²¹ consumed a high-cholesterol diet lacking added cholate for 12 weeks²². As in human atherosclerosis⁵, atheroma-associated cells in mice expressed CD40L and CD40

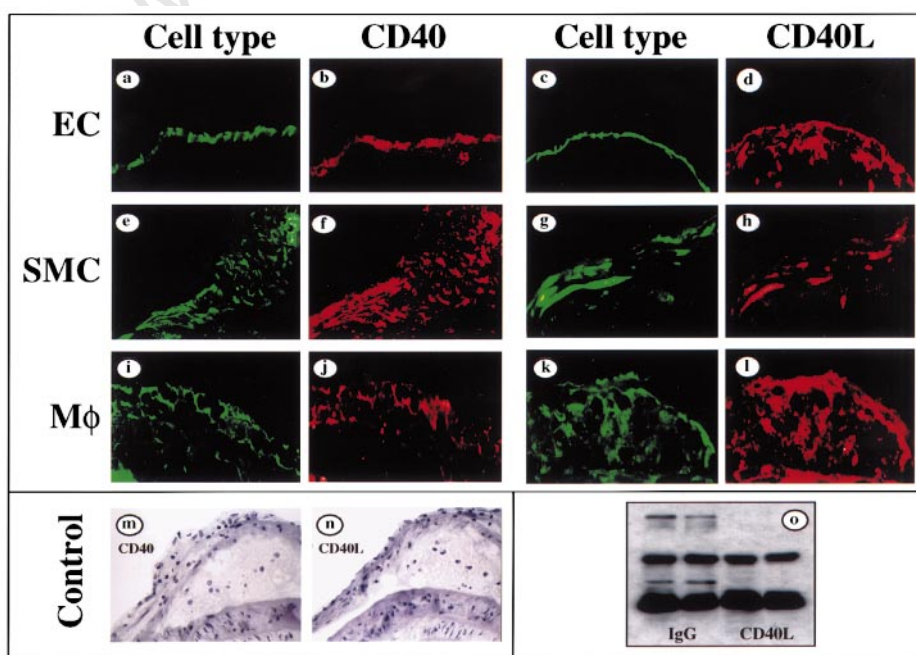


Figure 1 Expression of CD40 and CD40L in atherosclerotic lesions of LDL-receptor-deficient mice fed a high-cholesterol diet. Photomicrographs show frozen sections of aortic arches. The lumen of the artery is at the top of each picture. Immunofluorescence-double staining (for CD40 or CD40L, and a cell-type-specific marker) of frozen sections of aortic arches revealed expression of CD40 and CD40L on endothelial cells (EC) (a–d), smooth-muscle cells (SMC) (e–h), and macrophages (Mφ) (i–l). Controls with isotype immunoglobulin for

CD40 (m) and CD40L (n) are shown. Analysis of four different mice produced similar results. o, Detection of circulating rat anti-murine-CD40L antibody in mice. After 4, 8 and 12 weeks of administration of control antibody (IgG) or anti-CD40L antibody (CD40L), serum samples were prepared from all mice (3 days after the last injection), separated by SDS-PAGE and analysed by western blotting for the respective antibody. Two different serum analyses (after 12 weeks of experiments) are shown, representing each group.

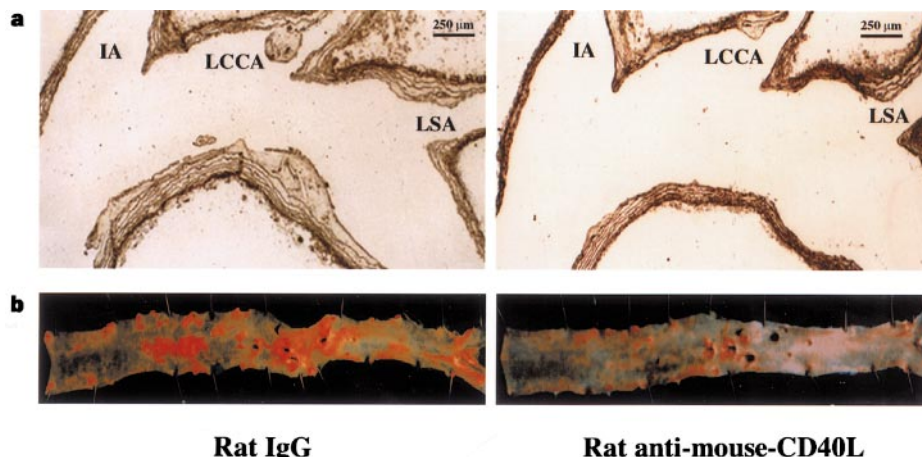


Figure 2 Reduction in aortic atherosclerosis by anti-CD40L treatment. Left panels show results of treatment with IgG and right panels results of treatment with anti-CD40L. **a**, The carotid arteries are at the top and the aortic sinus at the left of each picture. The ostia of the innominate artery (IA), the left common carotid artery

(LCCA) and the left subclavian artery (LSA) are indicated. **b**, The thoracic section of the aorta is at the left of each figure. Vessels were stained for lipid deposition. Representative specimens from both groups are shown.

(Fig. 1). Immunofluorescent double-staining with cell-type-specific antibodies showed that, within atherosclerotic lesions, CD40 and CD40L were expressed on endothelial cells, smooth-muscle cells and macrophages. CD40L was also expressed on CD4-positive T cells. To determine the role of CD40 signalling in the progression of atherogenesis, we randomly assigned mice to receive either a rat

anti-mouse-CD40L antibody ($n = 10$) or an irrelevant rat immunoglobulin (IgG) ($n = 8$) over a period of 12 weeks (both at $250 \mu\text{g}$ per mouse, twice a week by intraperitoneal injection). This regimen inhibits CD40L signalling *in vivo* by more than 90% (refs 23, 24). Indeed, after 4, 8 and 12 weeks, circulating rat IgG (in the groups treated with rat IgG or anti-CD40L monoclonal antibody) could be detected in the serum of treated mice after intraperitoneal administration of the antibody (Fig. 1o). A third group ($n = 6$) of LDL-receptor-deficient mice fed the atherogenic diet without any antibody treatment was a control for possible nonspecific effects of administration of rat IgG.

Anti-CD40L treatment significantly limited atherosclerotic lesion formation compared with either control group. Examination of the aortic arch (Fig. 2a), as well as of the abdominal aorta (Fig. 2b), showed a marked decrease in atherosclerotic lesion progression and aortic lipid deposition in mice treated with anti-CD40L antibody. Quantitative computer-assisted image analysis showed substantial reduction in the extent (59% reduction, $P < 0.002$) and thickness (41% reduction, $P < 0.001$) of aortic-arch lesions in the anti-CD40L-treated mice compared with rat-IgG-treated animals (Fig. 3a, b). For these analyses, we measured the full aortic wall to avoid any ambiguity in defining the intima. In addition, anti-CD40L antibody inhibited the formation of sudanophilic-lipid-rich lesions in abdominal aortas by 79% ($P < 0.0001$) (Fig. 3c). Furthermore, atheroma of mice treated with anti-CD40L antibody exhibited a 64% ($P < 0.002$) decrease in macrophage-specific staining (Fig. 4), contained 70% ($P < 0.001$) fewer T lymphocytes (Fig. 4), and showed markedly reduced expression of vascular cell adhesion molecule-1 (VCAM-1) in atherosclerotic lesions compared with controls (Fig. 5), concurring with *in vitro* studies that showed regulation of VCAM-1 on vascular cells through CD40 signalling^{17–19}.

Serum lipid profiles (total cholesterol, very-low-density lipoprotein, LDL and high-density lipoprotein; analysed by fast protein liquid chromatography²⁵), triglycerides, circulating leukocytes, haematocrit and body weight did not differ among the three groups. Furthermore, we did not detect antibodies to the rat anti-mouse-CD40L antibody in treated mice. Female LDL-receptor-deficient mice ($n = 8$), which developed ~30% fewer lesions than males as described^{26,27}, responded to anti-CD40L treatment with a decrease in atherosclerotic lesions that was similar to that of the treated males.

Our results establish a crucial regulatory role for CD40 signalling in atherogenesis in mice with raised LDL levels. Blocking CD40L–

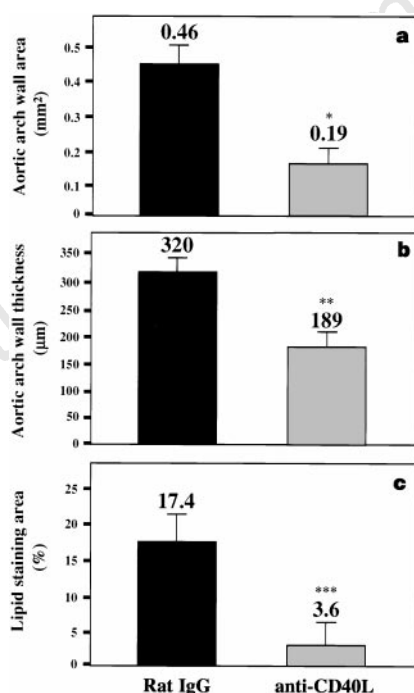


Figure 3 Reduction of atherosclerotic lesion size by anti-CD40L treatment. Photomicrographs of longitudinal sections of mice aorta were analysed by computer-assisted image quantification. Analysis of mice treated with irrelevant rat IgG (black bars) or anti-CD40L antibody (grey bars) is shown. **a**, **b**, The maximal wall area (**a**) and the maximal thickness (**b**) of the inner-aortic-arch wall for each mouse was measured as described in the Methods. **c**, The lipid-deposition area was quantified in abdominal aortas of all mice as described²⁷. The percentage of stain deposition was calculated as described in the Methods. Data represent mean values ($n = 10$ for anti-CD40L, $n = 8$ for rat IgG) with s.e.m. (single asterisk indicates $P < 0.002$; double asterisks indicate $P < 0.001$; and triple asterisks indicate $P < 0.0001$).

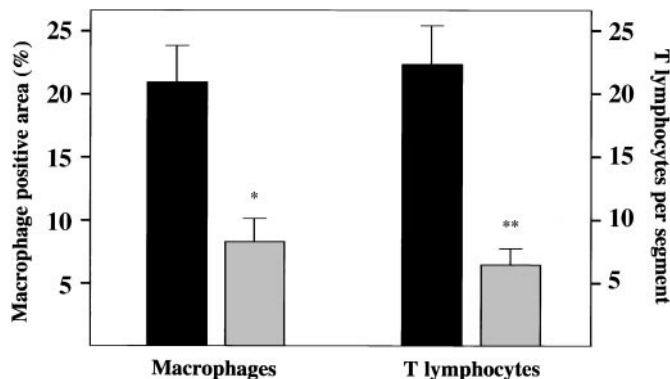


Figure 4 Reduction of macrophage and T-lymphocyte content in atheroma by blocking CD40 signalling. Photomicrographs of longitudinal sections of mice aorta were analysed by computer-assisted image quantification. The percentage of macrophage-positive areas and the total T-lymphocyte count were determined within the aortic-arch segment (see Methods) of mice treated with irrelevant rat IgG (black bars) and anti-CD40L antibody (grey bars). Data represent mean values ($n = 10$ for anti-CD40L, $n = 8$ for rat IgG) with s.e.m. (single asterisk indicates $P < 0.002$; and double asterisks indicate $P < 0.001$).

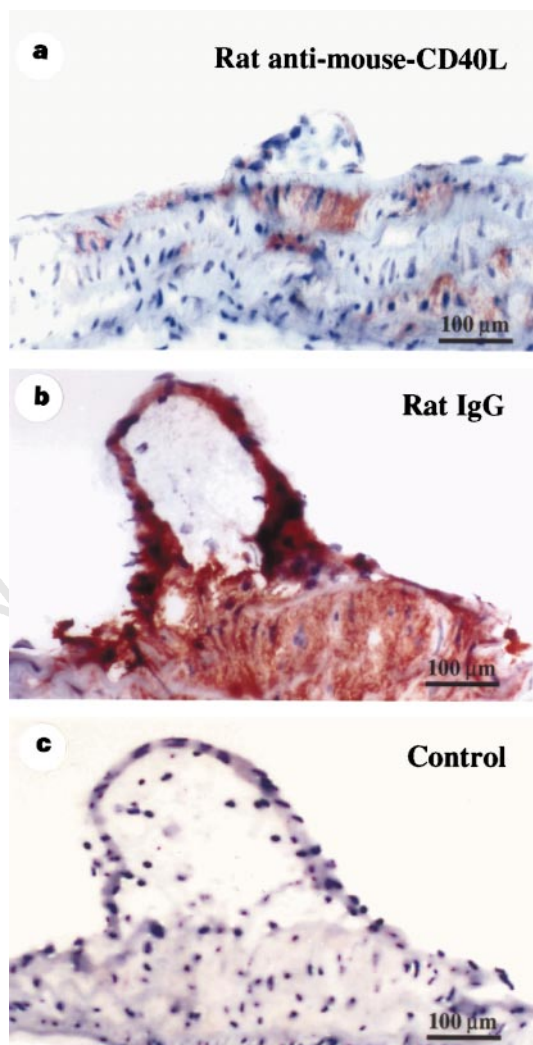


Figure 5 Reduction of VCAM-1 expression by blocking CD40 signalling. Photomicrographs show substantially reduced staining (red) for VCAM-1 in atherosclerotic lesions of anti-CD40L-treated mice (a) compared with lesions of rat-IgG-treated mice (b). For control, adjacent sections were stained with the isotype immunoglobulin (c). The lumen of the artery is at the top of each picture. Analysis of four different atheromas produced similar results.

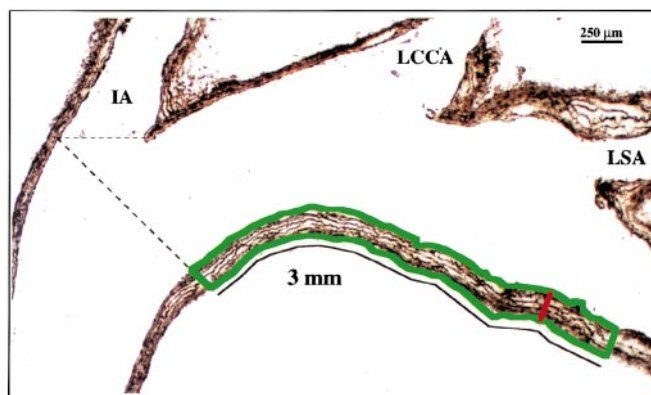


Figure 6 Measurements of aortic-arch-wall area and thickness in LDL-receptor-deficient mice. A longitudinal section of the aortic arch from an anti-CD40L-treated mouse is shown. Wall area (green line) and thickness (red line) were determined as described in the Methods. The ostia of the innominate artery (IA), the left common carotid artery (LCCA) and the left subclavian artery (LSA) are indicated.

CD40 signalling markedly decreased atherosclerotic plaque formation, arterial lipid deposition, and lesional expression of VCAM-1, a molecule considered important during atherogenesis. These findings are new evidence of the importance of inflammatory pathways in atheroma formation in response to elevated LDL, a common risk factor for coronary heart disease in humans.

Studies of compound mutant mice that lacked apolipoprotein E (apoE) and were globally immunodeficient because of interruption of recombination-activating gene-1 showed a 40% decrease in atherosclerotic lesions on a chow diet, but no change in lesion development when severe hypercholesterolaemia was induced by an atherogenic diet²⁸. These findings do not exclude a role for T lymphocytes in atherogenesis, particularly when the disease occurs without exaggerated hypercholesterolaemia, as is most often the case in human patients. Lesions in the LDL-receptor-deficient animals used here may have more T lymphocytes than these in the apoE-deficient mice²⁹. Although T lymphocytes are the traditional source of CD40L, other cell types within atheroma can also express this mediator⁵. Thus, interruption of CD40 signalling may reduce atheroma size independent of the involvement of T lymphocytes in atherogenesis.

CD40 signalling not only influences lesion evolution, as shown here, but also regulates several vascular-cell and macrophage functions related to the acute manifestations of atherosclerosis, including the thrombosis that ultimately precipitates unstable angina pectoris or acute myocardial infarction. CD40 ligation produces *in vitro* effects that are important in atherosclerosis and are not shared by 'classical' cytokines such as IL-1, TNF, or IFN- γ ; these effects include induction of procoagulant tissue factor in macrophages¹⁶. These findings, taken together, identify interruption of CD40 signalling as a new potential therapeutic target in atherogenesis. □

Methods

Treatment and analysis of mice. Eight-week-old LDL-receptor-deficient mice (Jackson Laboratory) consumed a high-cholesterol diet (Research Diets; 1.25% cholesterol, 0% cholate), and received either rat IgG ($n = 8$) or anti-CD40L antibody ($n = 10$) (Immunex Corporation) at 250 μ g per mouse by intraperitoneal injection twice a week. Anti-CD40L, a rat IgG2 monoclonal antibody raised against mouse CD40L, was prepared as described^{23,24}. After 12 weeks of treatment, mice were killed and aortas analysed. The arch and abdominal portions of the aorta were separated. The aortic arches were perfused with PBS and snap-frozen in optimal cutting temperature (OCT),

Tissue-Tek). To measure plaque size, longitudinal sections (about 30 per mouse) of the aortic arch were analysed microscopically for all mice. For this purpose, a 3-mm segment of the lesser curvature of the aortic arch was defined proximally by a perpendicular axis dropped from the right side of the innominate artery origin (Fig. 6, dashed line) and the aortic-arch wall area subtended by this 3-mm stretch of intima was calculated for each section of all mice by computerized image analysis (Fig. 6, green line). The aortic-arch wall thickness was determined on this same segment of the lesser curvature (Fig. 6, red line). The maximal area and thickness of the inner-aortic-arch wall of each mouse were used to compute averages per group.

For immunofluorescence double-staining, cryostat sections (5 µm) of the aortic arch were cut, air-dried, fixed in acetone, and stained with anti-CD40L, anti-CD40, or anti-VCAM-1 antibodies (PharMingen), or with the respective cell-specific antibody (anti-CD31 for endothelial cells, anti-α actin for smooth-muscle cells, anti-MAC3 for macrophages, and anti-CD3 for T lymphocytes; Dako) as described⁵. Areas that stained positive for macrophages were measured using computer-assisted image quantification (Optimas 5.2, Optimas Corporation). Lymphocytes identified by anti-CD3 staining were counted microscopically by two blinded observers. Controls for specificity included staining with the species-respective immunoglobulin (mouse myeloma protein MOPC-21, Sigma; rat and rabbit immunoglobulin, Dako). The abdominal aortas were fixed with 10% buffered formalin and stained for lipid deposition with Sudan IV (Fisher Scientific). The aortas were then opened longitudinally to the iliac bifurcation and pinned out on black wax surface using 0.2-mm steel pins. The percentage of stain deposition was calculated within the total surface area (15 mm from the iliac bifurcation to the thoracic section of the aorta) using computer analysis^{26,27}.

Received 9 February; accepted 27 April 1998.

- Ross, R. The pathogenesis of atherosclerosis: a perspective for the 1990s. *Nature* **362**, 801–809 (1993).
- Fuster, V., Lewis, A. Conner Memorial Lecture. Mechanisms leading to myocardial infarction: insights from studies of vascular biology. *Circulation* **90**, 2126–2146 (1996).
- Fuster, V., Badimon, L., Badimon, J. J. & Chesebro, J. H. The pathogenesis of coronary artery disease and the acute coronary syndromes. *N. Engl. J. Med.* **326**, 242–250 (1992).
- Hansson, G. K. et al. Localization of T lymphocytes and macrophages in fibrous and complicated human atherosclerotic plaques. *Arteriosclerosis* **72**, 135–141 (1988).
- Mach, F. et al. Functional CD40 ligand is expressed on human vascular endothelial cells, smooth muscle cells, and macrophages: implications for CD40-CD40 ligand signaling in atherosclerosis. *Proc. Natl Acad. Sci. USA* **94**, 1931–1936 (1997).
- Foy, T. M., Aruffo, A., Bajorath, J., Buhlmann, J. E. & Noelle, R. J. Immune regulation by CD40 and its ligand gp39. *Annu. Rev. Immunol.* **14**, 591–617 (1996).
- Zhou, X., Stemme, S. & Hansson, G. K. Evidence for a local immune response in atherosclerosis. *Am. J. Pathol.* **149**, 359–366 (1996).
- Lichtman, A. H., Cybulsky, M. & Lusinskas, F. W. Immunology of atherosclerosis: the promise of mouse models. *Am. J. Pathol.* **149**, 351–357 (1996).
- Emeson, E. E. & Robertson, A. L. Jr T lymphocytes in aortic and coronary intimas. Their potential role in atherogenesis. *Am. J. Pathol.* **130**, 369–375 (1988).
- Durie, F. H. et al. Prevention of collagen-induced arthritis with an antibody to gp39, the ligand for CD40. *Science* **261**, 1328–1330 (1993).
- Mohan, C., Shi, Y., Laman, J. D. & Datta, S. K. Interaction between CD40 and its ligand gp39 in the development of murine lupus nephritis. *J. Immunol.* **154**, 1470–1480 (1995).
- Larsen, C. P. et al. Long-term acceptance of skin and cardiac allografts after blocking CD40 and CD28 pathways. *Nature* **381**, 434–438 (1996).
- Gerritse, K. et al. CD40-CD40 ligand interactions in experimental allergic encephalomyelitis and multiple sclerosis. *Proc. Natl Acad. Sci. USA* **93**, 2499–2504 (1996).
- Carayanniotis, G., Master, S. R. & Noelle, R. J. Suppression of murine thyroiditis via blockade of the CD40-CD40L interaction. *Immunity* **90**, 421–426 (1997).
- Schönbeck, U. et al. Regulation of matrix metalloproteinase expression in human vascular smooth muscle cells by T lymphocytes: a role for CD40 signaling in plaque rupture? *Circ. Res.* **81**, 448–454 (1997).
- Mach, F., Schönbeck, U., Bonnefoy, J.-Y., Pober, J. S. & Libby, P. Activation of monocyte/macrophage functions related to acute atheroma complication by ligation of CD40. Induction of collagenase, stromelysin, and tissue factor. *Circulation* **96**, 396–399 (1997).
- Karmann, K., Hughes, C. W., Schechner, J., Fanslow, W. C. & Pober, J. S. CD40 on human endothelial cells: inducibility by cytokines and functional regulation of adhesion molecule expression. *Proc. Natl Acad. Sci. USA* **92**, 4342–4346 (1995).
- Yellin, M. J. et al. Functional interaction of T cells with endothelial cells: the role of D40L-CD40-mediated signals. *J. Exp. Med.* **182**, 1857–1864 (1995).
- Hollenbaugh, D. et al. Expression of functional CD40 by vascular endothelial cells. *J. Exp. Med.* **182**, 33–40 (1995).
- Mach, F., Schönbeck, U. & Libby, P. CD40 signaling in vascular cells: a key role in atherosclerosis? *Atherosclerosis* (in press).
- Ishibashi, S. et al. Hypercholesterolemia in low density lipoprotein receptor knockout mice and its reversal by adenovirus-mediated gene delivery. *J. Clin. Invest.* **92**, 883–893 (1993).
- Lichtman, A. H. et al. Comparative effects of precisely defined semipurified diets supplemented with lipid, cholesterol, and sodium cholate on serum-lipids and aortic atherosclerosis LDL. *FASEB J.* **11**, A154 (1997).
- Kennedy, M. K. et al. CD40/CD40 ligand interactions are required for T cell-dependent production of interleukin-12 by mouse macrophages. *Eur. J. Immunol.* **26**, 370–378 (1996).
- Lu, L. et al. Blockade of the CD40-CD40L pathway potentiates the capacity of donor-derived dendritic cell progenitors to induce long-term cardiac allograft survival. *Transplantation* **64**, 1808–1815 (1997).

- Yokode, Y., Hammer, R. E., Ishibashi, S., Brown, M. S. & Goldstein, J. L. Diet-induced hypercholesterolemia in mice: prevention by overexpression of LDL receptors. *Science* **250**, 1273–1275 (1990).
- Johnson, R. C. et al. Absence of P-selectin delays fatty streak formation in mice. *J. Clin. Invest.* **99**, 1037–1043 (1997).
- Tangirala, R. K., Rubin, E. M. & Palinski, W. Quantitation of atherosclerosis in murine models: correlation between lesions in the aortic origin and in the entire aorta, and differences in the extent of lesions between sexes in LDL receptor-deficient and apolipoprotein E-deficient mice. *J. Lipid Res.* **36**, 2320–2328 (1995).
- Dansky, H. M., Charlton, S. A., McGee Harper, M. & Smith, J. D. T and B lymphocytes play a minor role in atherosclerotic plaque formation in the apolipoprotein E-deficient mouse. *Proc. Natl Acad. Sci. USA* **94**, 4642–4646 (1997).
- Roselaar, S. E., Kakkannathu, P. X. & Daugherty, A. Lymphocyte populations in atherosclerotic lesions of ApoE^{−/−} and LDL receptor^{−/−} mice. Decreasing density with disease progression. *Arterioscler. Thromb. Vasc. Biol.* **16**, 1013–1018 (1996).

Acknowledgements. We thank E. Shvartz, E. Rabkin, and E. Simon-Morrissey for technical assistance; W. C. Fanslow III for providing the rat anti-mouse-CD40 antibody; and M. W. Freeman, K. J. Moore and V. Kunjathoor for assistance and discussions. This work was supported in part by grants from the NHLBI (to P. L.) and from the Fond National Suisse pour la Recherche Scientifique (to F. M.).

Correspondence and requests for materials should be addressed to P. L. (e-mail: plibby@rics.bwh.harvard.edu).

Truncating mutations of hSNF5/INI1 in aggressive paediatric cancer

Isabella Versteeg*, Nicolas Sévenet*, Julian Lange*, Marie-Françoise Rousseau-Merck*, Peter Ambros†, Rupert Handgretinger‡, Alain Aurias* & Olivier Delattre*

* Laboratoire de Pathologie Moléculaire des Cancers, Section de Recherche, Institut Curie, 26 rue d'Ulm, 75248 Paris Cedex 05, France

† CCRI, St Anna Kinderspital, Kinderspitalgasse 6, A-1090 Vienna, Austria

‡ Universität Kinderklinik, Rümelinstrasse 23, D-72070 Tübingen, Germany

Malignant rhabdoid tumours (MRTs) are extremely aggressive cancers of early childhood. They can occur in various locations, mainly the kidney, brain and soft tissues^{1,2}. Cytogenetic and molecular analyses have shown that the deletion of region 11.2 of the long arm of chromosome 22 (22q11.2) is a recurrent genetic characteristic of MRTs, indicating that this locus may encode a tumour suppressor gene^{3–8}. Here we map the most frequently deleted part of chromosome 22q11.2 from a panel of 13 MRT cell lines. We observed six homozygous deletions that delineate the smallest region of overlap between the cell lines. This region is found in the hSNF5/INI1 gene, which encodes a member of the chromatin-remodelling SWI/SNF multiprotein complexes^{9–12}. We analysed the sequence of hSNF5/INI1 and found frameshift or nonsense mutations of this gene in six other cell lines. These truncating mutations of one allele were associated with the loss of the other allele. Identical alterations were observed in corresponding primary tumour DNAs but not in matched constitutional DNAs, indicating that they had been acquired somatically. The observation of bi-allelic alterations of hSNF5/INI1 in MRTs suggests that loss-of-function mutations of hSNF5/INI1 contribute to oncogenesis.

To map in detail the homozygous deletions of chromosome band 22q11.2 that are observed in MRTs^{7,8}, we amplified 22q11.2 loci from MRT DNAs using the polymerase chain reaction (PCR). We used DNAs from established cell lines to avoid DNA contamination from non-tumour tissues. First, we tested 167 chromosome-22 loci from the DNA of a recently described MRT cell line⁸. For nine markers, specific PCR products were not detected in the tumour DNA, although the matched constitutional DNA promoted efficient amplification (cell line DL; Fig. 1). This suggested that the loci defined by these markers lay within the previously suspected homozygous deletion of 22q11.2 in DL⁸. These 9 markers and 49 flanking markers were then tested on the DNA from 12 other MRT cell lines. Loss of heterozygosity (LOH) was detected at polymorphic loci in three cases for which constitutional DNA was

LOCUS	TYPE	DL	G401	TM	LM	MON	LP	WT	MT	AS	2004	KD	Wa2	1783
3	D22S420	CA	0	-	-	0	●	0	0	-	-	0	0	-
	D22S427	CA	-	-	0	-	●	0	0	-	-	0	0	-
6	D22S264	CA	0	-	0	-	●	0	0	-	-	0	0	-
	D22S306	CA	-	-	0	-	●	0	0	-	-	0	0	-
0	D22S539	CA	●	-	0	0	●	●	●	-	-	-	0	-
	D22S446	CA	-	-	0	-	●	0	0	-	-	-	-	-
	D22S425	CA	●	-	0	-	●	●	●	-	-	-	0	-
	D22S257	CA	-	-	0	-	●	●	●	-	-	-	-	-
	D22S303	CA	●	-	0	-	●	●	●	-	-	-	-	-
	D22S301	CA	●	-	0	-	●	●	●	-	-	-	-	-
	A006E25	EST	■	■	■	■	■	-	-	-	-	-	-	-
	SGC32593	EST	■	■	■	■	■	-	-	-	-	-	-	-
2	MMP11	EST	■	■	■	■	■	-	-	-	-	-	-	-
	GCT10	TRI	■	■	■	■	■	-	-	-	-	-	-	-
	D22S345	CA	■	■	0	■	■	●	●	-	-	0	0	-
	D22S1106	EST	■	■	-	-	■	-	-	-	-	-	-	-
	SGC34185	EST	■	■	-	-	■	-	-	-	-	-	-	-
	D22S613	EST	■	■	-	-	■	●	●	-	-	-	-	-
	TOP1P2	CA	0	■	-	0	●	●	●	-	-	0	0	-
	D22S156	CA	-	-	-	0	●	●	●	-	-	0	0	-
	D22S419	CA	0	-	-	0	-	-	-	-	-	0	0	-
	D22S925	CA	0	-	-	0	-	-	-	-	-	0	0	-
0	D22S315	CA	-	-	-	0	-	●	●	-	-	0	0	-
	D22S1164	CA	0	-	-	0	●	●	●	-	-	0	0	-
	D22S1148	CA	0	-	-	0	●	●	●	-	-	0	0	-
	D22S926	CA	0	-	-	0	●	●	●	-	-	0	0	-
1	D22S421	CA	-	-	-	0	0	●	●	-	-	0	0	-
	D22S429	CA	-	-	-	-	●	●	●	-	-	-	-	-
11	D22S1154	CA	0	-	-	0	-	-	-	-	-	-	-	-
3	D22S1144	CA	0	-	0	0	●	●	●	-	-	0	0	-

Figure 1 The commonly deleted region in MRTs contains markers A006E25, SGC32593, MMP11 and GCT10. Results of the analysis of a subset of the markers tested by PCR amplification are indicated. At the left, the distances between markers are indicated in cM (centimorgans)²⁹. Filled squares indicate complete absence of amplification of a marker in the tumour DNA. Open and filled circles indicate retention and loss-of-heterozygosity of a marker sequence, respectively. A dash indicates the presence of a single PCR product in constitutional and/or tumour DNA. The type of marker is indicated: CA, dinucleotide polymorphic marker; TRI, trinucleotide polymorphic marker. The black bar at the right indicates the common region of deletion.

available (cell lines LP, WT and MT; Fig. 1), and highly probable in four other tumour DNAs with a single allele at numerous 22q11.2 polymorphic loci (cell lines AS, 2004, KD and Wa2; Fig. 1). We observed a complete absence of amplification of several loci in four additional cell lines (G401, TM, LM and MON; Fig. 1). From these results we could delineate a common region of deletion that includes markers A006E25, SGC32593, MMP11 and GCT10.

To eliminate the possibility that we were detecting artefacts of PCR, we confirmed that these loci were homozygously deleted by Southern blotting and hybridization of MRT DNAs with ³²P-labelled PCR products of the markers. This showed the complete absence, as compared with control DNAs, of the expected fragments in the DNAs from DL, G401, TM, LM and MON cell lines, an observation in agreement with the PCR results (data not shown). We used these markers to screen human bacteriophage P1-based artificial chromosome (PAC) and chromosome-22-specific cosmid libraries. The analysis of overlapping clones indicated that the common region of deletion spans less than 150 kilobases (kb) of DNA (data not shown).

A006E25 is an expressed sequence tag (EST) for *hSNF5/INI1*, a human homologue of the yeast *SNF5* gene^{9,10}. *hSNF5/INI1* encodes a member of the SWI/SNF complexes, which are thought to facilitate the transcriptional activation of inducible genes through the remodelling of chromatin¹¹⁻¹⁵. We focused on this gene because hybridization of a Southern blot of MRT DNAs with a 5'-*hSNF5/INI1* probe revealed an altered migration pattern of the germline 25-kb *Eco*R1 fragment of KD DNA (Fig. 2). To study this alteration further, we determined the genomic organization of the *hSNF5/INI1* gene on cosmid DNA by direct sequencing of exon/intron boundaries with primers derived from the complementary DNA sequence.

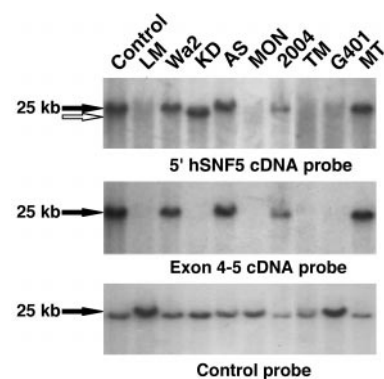


Figure 2 Homozygous deletion of exons 4 and 5 of *hSNF5/INI1* in the KD cell line. *Eco*R1-digested DNAs migrated on a 0.8% agarose gel were Southern-blotted and hybridized successively with a 5'-*hSNF5/INI1* cDNA probe (exon 1-6) (5'-*hSNF5*) and an exon 4-5 cDNA probe. Black arrows indicate the 25-kb germline DNA fragment and the white arrow shows the rearranged fragment in KD DNA. As a control for the quality of DNA in the 25-kb range, the same blot was hybridized with the SHGC-33468 EST probe.

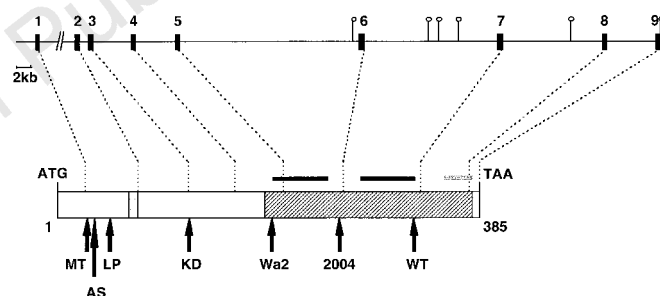


Figure 3 Genomic organization of the *hSNF5/INI1* gene and position of the stop codons generated by mutations of the coding sequence. The gene is shown at the top, and the protein at the bottom. Black boxes (top) indicate the position of exons (numbered 1-9). The localization of exons 5, 7 and 8 is not precisely known within respective *Eco*R1 sites (indicated by open circles). The hatched area represents the SNF5 domain. The grey box (bottom left) indicates a peptide sequence that can be present or absent in the protein, depending on the use of a cryptic splice donor site in exon 2. The black and grey bars above the *hSNF5* protein indicate the peptide repeats of the SNF5 domain and the possible C-terminal coiled-coil structure, respectively. Arrows indicate the position of stop codons generated by point mutations in the different tumours.

This indicated that the peptide sequence of the *hSNF5/INI1* gene is encoded by nine exons and that this gene extends over ~50 kb (Fig. 3). The amplification by reverse transcription with PCR (RT-PCR) and sequence of *hSNF5/INI1* messenger RNAs from the KD cell line showed an out-of-frame junction between exons 3 and 6. Furthermore, no PCR products were obtained for exons 4 and 5 using genomic DNA from this cell line as a template. The hybridization of the Southern blot of MRT DNAs with an *hSNF5/INI1* cDNA probe consisting of exons 4 and 5 detected the expected 25-kb fragments from control DNAs, but no signal from the KD DNA, showing that the altered migration pattern of this fragment was linked to a genomic deletion that includes these two exons (Fig. 2). These results showed that the smallest region of homozygous deletion observed in MRTs lay within the *hSNF5/INI1* gene.

We searched for mutations of the *hSNF5/INI1* gene in the seven non-homozygously deleted cell lines by direct sequencing of both RT-PCR-amplified *hSNF5/INI1* mRNAs and PCR products of individual exons obtained with primers derived from intron sequences. In six cases, DNA variants were identified (Table 1).

Table 1 Analysis of the *hSNF5/INI1* gene in cell lines, primary tumours and constitutional DNAs from MRT patients

Name of cell line	Age*	Localization	Cytogenetics	Analysis of RNA and DNA			
				Cell lines		Primary	
				Allele 1	Allele 2	tumour†	Constitutional
DL	14	Lung	46XY,t(1;22)	Deletion	Deletion	–	Two alleles
LP	12 years	Kidney	46XY	47TAC/TAA	Deletion	id	Two alleles, NS
WT	21	Kidney	46XY,-15,der15 t(1;15)	317, del 1 bp	Deletion	id	Two alleles, NS
MT	1	Soft tissue	46XY	31, ins 72 bp‡	Deletion	id	Two alleles, NS
G401	3	Kidney	46XY	Deletion	Deletion	–	–
MON	6	Abdomen	46XY,t(7;22)	Deletion	Deletion	–	Two alleles
TM	21	Retroperitoneal	46XY,t(11;22)	Deletion	Deletion	–	–
2004	6	Kidney	46XY	258, del 13 bp, ins 2 bp	Deletion	id	Two alleles, NS
1783	6	Kidney	46XY	NAD	NAD	–	–
KD	11	Abdomen	46XX, 9p+	Deletion of exons 4 and 5	Deletion	–	–
LM	8	Liver	46XX,3p+, 22q–	Deletion	Deletion	–	–
Wa2	17	Intraspinal	46XY,t(18;22)	196, dup 17 bp	Deletion	–	–
AS	7	Abdomen	46XX	37, del 19 bp	Deletion	–	–

*Age of the patient is shown in months except for the LP cell line. † Tumour material is derived from surgical biopsy. ‡ This insertion contains an in-frame stop codon. Deletion indicates a complete deletion of the *hSNF5/INI1* gene; NAD, no alteration detected; dash, no material available for analysis; id, identical alterations observed in the primary tumour and in the cell line; NS, normal sequence. For each point mutation, the position of the codon and the type of alteration are indicated; bp, base pairs; del, deletion; ins, insertion; dup, duplication.

Frameshifts due to nucleotide deletions and/or insertions were detected in five cases and a TAC-to-TAA nonsense mutation was observed in one case (Table 1). These alterations were observed in both cDNAs and corresponding genomic sequences. Double sequences, which might have been indicative of the presence of a wild-type *hSNF5/INI1* allele, were not detected. This observation confirmed LOH analyses and showed the alteration of both alleles in these six cases.

To exclude *in vitro* growth as a cause of these mutations, we analysed DNAs isolated from the matched primary tumours. Identical mutations were observed in all tested cases, establishing their presence in the original tumour cells (Table 1). Matched constitutional DNA was available in six cases. For four cases in which precise sequence alterations were found in tumour DNA (tumours LP, WT, MT and 2004), matched constitutional DNAs contained wild-type sequences. In two tumours with homozygous deletions (DL and MON), constitutional DNAs exhibited heterozygosity at the GCT10 locus (Table 1). Finally, neither LOH nor aberrant sequences could be detected in the 1783 tumour DNA. This may indicate that subtle mutations affecting non-coding sequences, including the promoter region, may have occurred in this tumour. Alternatively, it may indicate that, in rare cases of MRT, *hSNF5/INI1* may not be altered, or that, the diagnosis of MRT frequently being difficult, 1783 may be a different pathological entity to MRT.

These data show that bi-allelic, somatic alteration of *hSNF5/INI1*, arising through a combination of complete deletions of the gene, frameshift alterations and nonsense mutations, is a frequent, if not constant, feature of MRTs. This is fully consistent with the paradigm of the 'two-hit' recessive model of oncogenesis and supports the hypothesis that *hSNF5/INI1* is the MRT tumour suppressor gene. Although constitutional mutations were not detected in this study, which was based on sporadic tumours, previous reports of multifocal and/or familial MRTs indicate that constitutional inactivation of one allele of *hSNF5/INI1* could be responsible for hereditary cases^{5,16}. Given the low incidence of MRTs and its aggressive evolution, which leads to frequent death at a young age, familial cases, if existing, are expected to be rare. Most of the mutations described here are predicted to truncate the phylogenetically conserved 200-amino-acid domain that mediates interactions with hbrm and with the HIV integrase^{9,17}. However, the mutation in the WT cell line mainly results in the truncation of the 68 carboxy-terminal amino acids that probably form a coiled-coil structure, suggesting that the impairment of *hSNF5/INI1*-specific interactions, mediated by this C-terminal domain, is essential for progression of MRTs (Fig. 3).

The alteration of chromosome 22 was the only recurrent cytogenetic abnormality found in this series of tumours (Table 1)^{3–8}.

These observations and the high frequency of truncating mutations of *hSNF5/INI1* that we observed indicate that alteration of this gene is a major oncogenic event in MRT and could thus recapitulate several characteristics of this highly malignant paediatric cancer, including abnormal proliferation, local invasiveness and high propensity to undergo metastasis.

Accumulating evidence indicates that altered chromatin organization at specific DNA sites might be crucial in the process of oncogenesis. Indeed, the regulation of post-translational modifications of histones has been shown to be the target of alteration in cancer^{18–22}. The SWI/SNF complexes, which have been identified in organisms from yeasts to humans, are also thought to be important in the remodelling of chromatin structure. Members of these complexes have also been linked to the control of cell proliferation, in particular by binding to the retinoblastoma protein^{23–25}. Our results are, to our knowledge, the first to show that a member of the SWI/SNF complexes has characteristics of a tumour suppressor. The results indicate that the transcription of key genes involved in growth, differentiation or apoptotic processes could be selectively regulated by these complexes. □

Methods

Patients' samples. The TM87-16 (TM), G401, DL, 2001 and Wa2 cell lines have been described^{7,8,26,27}. Other cell lines were established through *in vitro* growth of MRT cells in RPMI 1640 medium plus 10% fetal calf serum. Part of the primary tumour samples was directly frozen in liquid nitrogen. Histopathological analysis of all cases showed that the cells were round to polygonal in shape, with vesicular nuclei, prominent nucleoli, eosinophilic cytoplasm and rare but typical cytoplasmic inclusions. The cells stained positive for both keratin and vimentin and negative for desmin. Constitutional material was obtained from blood or non-tumour tissues. DNAs and RNAs were isolated using standard procedures.

PCR. The sequences of the primers were retrieved from the Genome Database (mirror site at <http://gdb.infobiogen.fr/>), the Whitehead Institute/MIT genome center (<http://www-genome.wi.mit.edu/>), the Human Gene Map (<http://www.ncbi.nlm.nih.gov/SCIENCE96/>), the Cooperative Human Linkage Center (<http://www.chlc.org:80/>), the Stanford Human Genome Center (<http://www-shgc.stanford.edu>) and the Sanger Center (<http://www.sanger.ac.uk/>). PCR was performed using the GeneAmp PCR kit (Perkin Elmer) according to the programmes provided by the above centres.

PACs and cosmid screening and sequencing. Filters from a gridded human PAC library (RPC11.3-5 from the Lawrence Livermore National Laboratory)²⁸ provided by the Ressourcen Zentrum Primardaten Bank, and from a chromosome-22 library (LL22NC03), were screened with labelled PCR products from the *hSNF5/INI1* cDNA. After amplification of the positive clones, DNA was isolated using the Qiagen Plasmid kit and used to determine the *EcoRI* restriction map of the *hSNF5/INI1* locus. Sequences of exon/intron boundaries

were obtained through direct sequencing of DNAs from overlapping cosmids with primers derived from the cDNA sequence using the AmpliTaqFS BigDye Terminator kit (Applied Biosystems). Sequencing reactions were resolved on an ABI 377 Prism automated sequencer.

Southern blot analysis. 10 µg EcoRI-digested cell-line and constitutional DNAs from MRT patients and from non-affected controls were subjected to electrophoresis in 0.8% agarose gel, transferred to nylon membranes (Hybond N+) and successively hybridized with ³²P-labelled PCR products from A006E25, MMP11 and SGC32593 DNAs, 5'-*hSNF5/INI1* cDNA (exon 1-6), 3'-*hSNF5/INI1* cDNA (exon 6-9), exons 4-5 *hSNF5/INI1* cDNA and the SHGC-33468 EST probes.

Mutation detection. We completely analysed the coding sequences of both strands of both the *hSNF5/INI1* cDNAs and the genomic DNAs from each cell line. RNAs were reverse-transcribed with oligo-dT primers and resulting cDNAs were amplified using the GeneAmp RNA PCR core kit (Perkin Elmer). In each case, two overlapping PCR products corresponding to 5'-*hSNF5/INI1* and 3'-*hSNF5/INI1* cDNA were analysed. For the analysis of genomic DNA, we designed primer couples located in introns and used these primers to PCR-amplify the nine exons of *hSNF5/INI1* individually. All PCR amplification reactions were performed using the GeneAmp RNA PCR kit (Perkin Elmer) in 1.5 mM MgCl₂ with the following parameters: denaturation, 94°C for 30 s; annealing, 60°C for 30 s; and extension, 72°C for 30 s. PCR products were directly sequenced using the AmpliTaqFS Dye or BigDye Terminator kits (Applied Biosystems) on automatic 373A or 377 Prism sequencers. Primer sequences are available on request.

Received 26 February; accepted 7 May 1998.

- Parham, D. M., Weeks, D. A. & Beckwith, J. B. The clinicopathologic spectrum of putative extrarenal rhabdoid tumours. An analysis of 42 cases studied with immunohistochemistry or electron microscopy. *Am. J. Surg. Pathol.* **18**, 1010-1029 (1994).
- Wick, M. R., Ritter, J. H. & Dehner, L. P. Malignant rhabdoid tumours: a clinicopathologic review and conceptual discussion. *Semin. Diagn. Pathol.* **12**, 233-248 (1995).
- Douglass, E. C. *et al.* Malignant rhabdoid tumour: a highly malignant childhood tumour with minimal karyotypic changes. *Gene Chromosom. Cancer* **2**, 210-216 (1990).
- Shashi, V., Lovell, M. A., von Kap-herr, C., Waldron, P. & Golden, W. L. Malignant rhabdoid tumour of the kidney: involvement of chromosome 22. *Gene Chromosom. Cancer* **10**, 49-54 (1994).
- Fort, D. W., Tonk, V. S., Tomlinson, G. E., Timmons, C. F. & Schneider, N. R. Rhabdoid tumour of the kidney with primitive neuroectodermal tumour of the central nervous system: associated tumours with different histologic, cytogenetic, and molecular findings. *Gene Chromosom. Cancer* **11**, 146-152 (1994).
- Schofield, D. E., Beckwith, J. B. & Sklar, J. Loss of heterozygosity at chromosome regions 22q11-12 and 11p15.5 in renal rhabdoid tumours. *Gene Chromosom. Cancer* **15**, 10-17 (1996).
- Biegel, J. A. *et al.* Narrowing the critical region for a rhabdoid tumour locus in 22q11. *Gene Chromosom. Cancer* **16**, 94-105 (1996).
- Rosty, C. *et al.* Cytogenetic and molecular analysis of a t(1;22)(p36;q11.2) in a rhabdoid tumour with a putative homozygous deletion of chromosome 22. *Gene Chromosom. Cancer* **21**, 82-89 (1998).

- Muchardt, C., Sardet, C., Bourachot, B., Onufryk, C. & Yaniv, M. A human protein with homology to *Saccharomyces cerevisiae* SNF5 interacts with the potential helicase hbrm. *Nucleic Acid. Res.* **23**, 1127-1132 (1995).
- Kalpana, G. V., Marmion, S., Wang, W., Crabtree, G. R. & Goff, S. P. Binding and stimulation of HIV-1 integrase by a human homolog of yeast transcription factor SNF5. *Science* **266**, 2002-2006 (1994).
- Peterson, C. L. Multiple SWItches to turn on chromatin? *Curr. Opin. Genet. Dev.* **6**, 171-175 (1996).
- Wang, W. *et al.* Diversity and specialisation of mammalian SWI/SNF complexes. *Genes Dev.* **10**, 2117-2130 (1996).
- Kwon, H., Imbalzano, A. N., Khavari, P. A., Kingston, R. E. & Green, M. R. Nucleosome disruption and enhancement of activator binding by a human SWI/SNF complex. *Nature* **370**, 477-485 (1994).
- Owen-Hughes, T., Utley, R. T., Côté, J., Peterson, C. L. & Workman, J. L. Persistent site-specific remodeling of a nucleosome array by transient action of the SWI/SNF complex. *Science* **273**, 513-516 (1996).
- Côté, J., Quinn, J., Workman, J. L. & Peterson, C. L. Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. *Science* **265**, 53-60 (1994).
- Lynch, H. T. *et al.* Paravertebral malignant rhabdoid tumour in infancy. *In vitro* studies of a familial tumour. *Cancer* **52**, 290-296 (1983).
- Morozov, A., Yung, E. & Kalpana, G. V. Structure-function analysis of integrase interactor 1/hSNF5L1 reveals differential properties of two repeat motifs present in the highly conserved region. *Proc. Natl Acad. Sci. USA* **95**, 1120-1125 (1998).
- Sobulo, O. M. *et al.* MLL is fused to CBP, a histone acetyltransferase, in therapy-related acute myeloid leukemia with a t(11;16)(q23;p13.3). *Proc. Natl Acad. Sci. USA* **94**, 8732-8737 (1997).
- Magnaghi-Jaulin, L. *et al.* Retinoblastoma protein represses transcription by recruiting a histone deacetylase. *Nature* **391**, 601-604 (1998).
- Brehm, A. *et al.* Retinoblastoma protein recruits histone deacetylase to repress transcription. *Nature* **391**, 597-601 (1998).
- Lin, R. J. *et al.* Role of histone deacetylase complex in acute promyelocytic leukaemia. *Nature* **391**, 811-814 (1998).
- Grignani, F. Fusion proteins of the retinoic acid receptor-α recruit histone deacetylase in promyelocytic leukaemia. *Nature* **391**, 815-818 (1998).
- Dunaief, J. L. *et al.* The retinoblastoma protein and BRG1 form a complex and cooperate to induce cell cycle arrest. *Cell* **79**, 119-130 (1994).
- Trouche, D., Le Chalony, C., Muchardt, C., Yaniv, M. & Kouzarides, T. RB and hbrm cooperate to repress the activation functions of E2F1. *Proc. Natl Acad. Sci. USA* **94**, 11268-11273 (1997).
- Muchardt, C., Bourachot, B., Reyes, J.-C. & Yaniv, M. *ras* transformation is associated with decreased expression of the brm/SNF2α ATPase from the mammalian SWI-SNF complex. *EMBO J.* **16**, 101-109 (1998).
- Soulié, J., Rousseau-Merck, M.-F., Mouly, H. & Nezelof, C. Cytogenetic study of cell lines from an infantile hypercalcemic renal tumour. *Cancer Genet. Cytogenet.* **21**, 117-122 (1985).
- Handgretinger, R. *et al.* Establishment and characterisation of a cell line (Wa-2) derived from an extrarenal rhabdoid tumour. *Cancer Res.* **50**, 2177-2182 (1990).
- Iannou, P. *et al.* A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nature Genet.* **6**, 84-89 (1994).
- Dib, C. *et al.* A comprehensive genetic map of the human genome based on 5,264 microsatellites. *Nature* **380**, 152-154 (1996).

Acknowledgements. We thank M. Yaniv and C. Muchardt for discussions; T. Melot, A. Laugé, S. Pagès, M. Peter, I. Legrand, C. Rosty, J. Couturier and D. Stoppa-Lyonnet for their help; and the following clinicians for providing samples used in this study: F. Doz, J. Michon, H. Pacquement, E. Quintana, P. Ruck and J.-M. Zucker. I.V. and N.S. are recipients of fellowships from the European Union and the Ministère de l'Éducation Nationale, de la Recherche et de la Technologie, respectively. This work was supported by grants from the Association pour la Recherche contre le Cancer, the Ligue Nationale Contre le Cancer, the Institut Curie and the Programme Hospitalier de Recherche Clinique.

Correspondence and requests for materials should be addressed to O.D. (e-mail: delattre@curie.fr). EMBL accession numbers for intron-exon boundaries of *hSNF5/INI1* are Y17118-Y17126.

KNOW YOUR COPY RIGHTS RESPECT OURS

The publication you are reading is protected by copyright law.
Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA.

Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited
90 Tottenham Court Road, London W1P 0LP
Telephone: 0171 436 5931 Fax: 0171 436 3986